THE UNIVERSITY
of EDINBURGH

Explicit numerical algorithms for stochastic differential equations and their applications in data science and optimization

Ying Zhang

Doctor of Philosophy
University of Edinburgh
2020

Lay Summary

A wide range of real-world problems, especially in data science and finance, can be described by stochastic differential equations (SDEs). However, there are usually no analytical forms to the solutions of the SDEs, and it is natural to seek numerical approximations. This thesis is concerned with different explicit numerical algorithms which are of high computational efficiency compared to implicit methods. Under relaxed conditions, these explicit algorithms are examined in the problem of sampling and stochastic optimization, which are popular applications from Bayesian statistics, machine learning, and finance. Theoretical guarantees for the convergence properties of these algorithms are provided by using techniques from stochastic analysis, PDE theory, statistics, and financial mathematics.

Abstract

In this thesis, we focus on the analysis of a new generation of explicit numerical algorithms for (non-linear) random systems and their applications in computational statistics, non-convex optimization, data science, and finance.

In the first part, a tamed (explicit) order 1.5 algorithm is proposed to approximate solutions of stochastic differential equations (SDEs) with super-linear coefficients. Under certain conditions on the coefficients, a convergence result in $\mathscr{L}^2$ is obtained.

Then, in the second part, a new higher order Langevin Monte Carlo (LMC) algorithm is constructed, which is based on the application of an order 1.5 scheme to the Langevin SDE. The proposed LMC algorithm is considered in the context of sampling from a target distribution π that has a density $\hat{\pi}$ on $\mathbb{R}^d$ known up to a normalizing constant. To obtain the convergence results, $-\log \hat{\pi}$ is assumed to have a locally Lipschitz gradient and its third derivative is locally Hölder continuous with exponent $\alpha \in (0, 1]$. Non-asymptotic bounds are established for the convergence to stationarity of the new sampling method with convergence rate $1 + \alpha/2$ in Wasserstein distance, while it is shown that the rate is 1 in total variation even in the absence of convexity. In particular, in the case where $-\log \hat{\pi}$ is strongly convex and its gradient is Lipschitz continuous, explicit constants are provided.

Finally, in the last part of the thesis, the stochastic gradient Langevin dynamics (SGLD) algorithm is considered, which can be viewed as an extension of the unadjusted Langevin algorithm (ULA). We focus on a stochastic optimization problem via the SGLD algorithm. Crucially, the aim is to obtain non-asymptotic error bounds of the SGLD algorithm with discontinuous gradient $H(\theta, x)$. Theoretical guarantees in Wasserstein distances are provided under such an assumption for both convex and non-convex objective functions. Moreover, explicit upper estimates are obtained for the expected excess risk associated with the approximation of global minimizers of these objective functions. Numerical results for key examples in statistics and finance are presented which support the main findings.

Acknowledgements

First of all, I would like to express my gratitude to my supervisor, Dr. Sotirios Sabanis, for his support, patience and encouragement throughout my PhD study. He taught me a lot of mathematics, and provided me with many opportunities to work on collaborative projects and to present my research at different workshops. He always encouraged me when I felt stressed, and his enthusiasm for research influenced me deeply. I am also grateful to Dr. Miklos Rasonyi and Prof. Eric Moulines for their helpful advice and guidance on joint projects that lead to the work included in this thesis.

I would like to thank Prof. Tony Carbery and Prof. Dugald Duncan for giving me the opportunity to be part of the MIGSAA programme. I would also like to thank Ms. Isabelle Hanlon for her constant support in administrative works.

I am grateful to Sizhou, Guopeng, Xiling, Wenyi, and Shuren for the food, chat and all the fun we had together. I also thank Stefania, William, Lukas, Felix, Christina, and other colleagues for making my experience at Edinburgh special and pleasant.

Finally, I am thankful to my mom for her support throughout the period of my study. She is always my motivation for everything in my life.

To my mother

Contents

Chapter 1

Introduction

1.1 Literature review and brief summary of results

Stochastic differential equations (SDEs) are widely used to model real-world phenomena. They have found applications in various fields such as economics and finance, data science, engineering, physics, chemistry and biology. Nevertheless, SDEs usually do not admit explicit solutions, which makes it necessary to develop numerical schemes to approximate the solutions of SDEs.

For SDEs with globally-Lipschitz coefficients, the explicit Euler method, see e.g. [27] and [36], is commonly used to obtain approximations of the solutions, as it is easily implementable and of high computational efficiency compared to implicit methods. However, for SDEs with super-linearly growing coefficients, it is shown in [24, Theorem 1] that the absolute moments of the explicit Euler approximations at a finite time could diverge to infinity, which implies, in this case, one can not obtain strong convergence results of the explicit Euler scheme. To cope with this problem, a tamed Euler scheme is introduced in [25], where the drift term is modified using a specified taming factor. The (aforementioned) tamed scheme is applied to approximate the solution of an SDE that has a globally Lipschitz diffusion coefficient, and a globally one-sided Lipschitz drift coefficient whose derivative grows at most polynomially. [25, Theorem 1.1] shows that the tamed Euler approximations converge strongly with rate 1/2 in $\mathscr{L}^p$ to the exact solution of the SDE. Then, in [44] and [45], the tamed Euler scheme is improved and a new class of explicit Euler schemes is proposed to approximate SDEs with super-linearly growing drift and diffusion coefficients. $\mathscr{L}^p$ convergence results of such explicit Euler-type schemes are presented in [45, Theorem 1] under mild conditions. Moreover, the taming technique has been extended to Milstein-type schemes, see [5], [50], [30] and references therein. It is proved that the tamed Milstein scheme converges strongly in $\mathscr{L}^p$ with an improved rate 1 compared to the tamed Euler scheme.

By studying the results on the tamed Euler and Milstein schemes, one naturally thinks of the possibility to extend the taming technique to higher order schemes. A conjecture appears in [30, Remark 2], where it is stated that it is possible to construct, in a specified way, any high order explicit numerical schemes to approximate the solutions of SDEs with superlinear coefficients. Motivated by this conjecture, in Chapter 2, a tamed order 1.5 scheme is proposed. One notes that the aforementioned scheme is a "tamed" version of the order 1.5 scheme introduced in [27, Chapter 10.4], which is constructed based on the Itô-Taylor (known also as Wagner-Platen) expansion, see [27, Chapter 5.5] and [41]. In order to establish the strong convergence result, the coefficients of the super-linear SDE are assumed to satisfy suitable growth and globally one-sided Lipschitz conditions, while their second derivatives are Hölder continuous with exponent $\alpha \in (0, 1]$. Under these conditions, Theorem 2.6 shows that the explicit tamed order 1.5 scheme converges strongly in $\mathscr{L}^2$ with rate $1 + \alpha/2$. This answers the conjecture to the positive for the case of order 1.5 approximations. Furthermore, by the combination of the results in [45], [30] and in Chapter 2, one can arguably anticipate that, by using the uniform taming approach as explained in Section 2.2, any high order (explicit) scheme can be constructed with the desired rate of convergence as in the global Lipschitz case (see [41]).

Recent developments in data science attracted our attention to the fact that high order schemes can be applied to MCMC algorithms with improved convergence properties in high dimensions. This motivates the use of the tamed order 1.5 scheme in data-driven applications. In particular, the sampling problem in Bayesian statistics and machine learning is considered, which can be formulated as follows: we want to sample efficiently from a high-dimensional target probability distribution π that has a density on $\mathbb{R}^d$, denoted by $\hat{\pi}$, such that

$$\hat{\pi}(x) = e^{-U(x)} / \int_{\mathbb{R}^d} e^{-U(y)} \, dy,$$

with $\int_{\mathbb{R}^d} e^{-U(y)} \, dy < \infty$, where $U : \mathbb{R}^d \to \mathbb{R}$ is typically continuously differentiable. Within such a setting, consider a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$, then the Langevin SDE associated with π is defined by

$$dx_t = -\nabla U(x_t)dt + \sqrt{2}dw_t, \tag{1.1}$$

where $(w_t)_{t\geq 0}$ is a d-dimensional Brownian motion. It is a classical result that under mild conditions, the SDE (1.1) admits π as its unique invariant measure. Thus, to sample from the target distribution π, a commonly used method is to discretize the Langevin SDE (1.1) using the Euler-Maruyama (Milstein) scheme, which yields the unadjusted Langevin algorithm (ULA), known also as the Langevin Monte Carlo (LMC) algorithm. The sampling behaviour of the ULA algorithm has been well studied in the literature and theoretical guarantees are established. For a globally Lipschitz ∇U, the non-asymptotic bounds in total variation and Wasserstein distance between the n-th iteration of the ULA algorithm and π have been provided in [11], [15] and [14]. As for the case of superlinear ∇U, again, the difficulty arises from the fact that the algorithms constructed based on explicit numerical schemes, for example ULA, is unstable (see [37]), and its Metropolis adjusted version, MALA, loses some of its appealing properties as discussed in [43] and demonstrated numerically in [7]. However, this problem can be addressed by using the taming technique, and the tamed unadjusted Langevin algorithm (TULA) is proposed in [7]. Moreover, by extending the strong convergence results of the tamed explicit schemes in [5], [25], [30], [44], [45], [50], non-asymptotic bounds of the TULA algorithm are obtained in total variation and in Wasserstein distance in [7, Theorem 4, 5].

In Chapter 3, a new higher order LMC algorithm (HOLA) is constructed, which is obtained by applying the order 1.5 scheme in [27] to the Langevin SDE (1.1). We consider the case where ∇U is super-linearly growing, and thus the taming technique is used. However, one notes that, unlike the tamed order 1.5 scheme discussed in Chapter 2, it is not suitable to apply a uniform taming approach suggested in [30, Remark 2] to the HOLA algorithm. Instead, different taming factors are used for different terms of the HOLA algorithm in order to obtain exponential moment bounds via the Log-Sobolev inequality, see Proposition 3.10, and consequently, convergence results can be established in an infinite time horizon. Under certain conditions (specified in Section 3.2), it can be shown that the proposed algorithm has a unique invariant measure π_γ, and moreover, Theorem 3.4 and 3.5 state that the rate of convergence between the law of the n-th iteration of the HOLA algorithm and the target measure π is $1 + \alpha/2$ in Wasserstein-2 distance, while the rate is 1 in total variation.

However, for the sampling problem described above, the gradient of $-\log \hat{\pi}$, i.e. ∇U, is usually unknown and one only has an unbiased estimate of ∇U. A natural extension of the ULA (LMC) algorithm, which was introduced in [51], is the stochastic gradient Langevin dynamics (SGLD) algorithm. We consider the application of the SGLD algorithm (1.3) in a stochastic optimization problem:

$$\text{minimize} \quad \hat{U}(\theta) := \mathbb{E}[f(\theta, Z)],$$

where $\theta \in \mathbb{R}^d$ and Z is a random element. We aim to generate $\hat{\theta}$ such that the expected excess risk

$$\mathbb{E}[\hat{U}(\hat{\theta})] - \inf_{\theta \in \mathbb{R}^d} \hat{U}(\theta)$$

is minimized. One notices that the optimization problem is linked to the problem of sampling from the target distribution $\pi_\beta \propto \exp(-\beta \hat{U}(\theta))$. This is due to the fact that π_β concentrates around the minimizers of $\hat{U}$ when β is sufficiently large, see [26]. For the sampling problem, it is well-known that under mild conditions, the Langevin SDE associated with π_β is given by

$$d\hat{Y}_t = -h(\hat{Y}_t)dt + \sqrt{2\beta^{-1}}dw_t, \tag{1.2}$$

where $h(\theta) = \nabla \hat{U}$ and $(w_t)_{t\geq 0}$ represents the standard Brownian motion. Then, to sample from π_β, the SGLD algorithm of the SDE (1.2) can be expressed as, for any $n \in \mathbb{N}$,

$$\theta_{n+1}^\gamma = \theta_n^\gamma - \gamma H(\theta_n^\gamma, Z_{n+1}) + \sqrt{2\beta^{-1}\gamma}\xi_{n+1}, \quad \theta_0^\gamma = \theta_0, \tag{1.3}$$

where θ_0 is an $\mathbb{R}^d$-valued random variable, $\gamma > 0$ is the stepsize, $\beta > 0$ is the so-called inverse temperature parameter, $H : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ is a measurable function satisfying $h(\theta) = \mathbb{E}[H(\theta, Z_0)]$ with $(Z_n)_{n\in\mathbb{N}}$ being a sequence of i.i.d. random variables, and $(\xi_n)_{n\in\mathbb{N}}$ is a sequence of standard independent d-dimensional Gaussian random variables. Our goal is thus to show that the law of the SGLD algorithm (1.3) is close to π_β in some proper sense, and therefore, the output of the SGLD algorithm (1.3) for large n is an almost-minimizer of $\hat{U}$.

Theoretical guarantees for the SGLD algorithm (1.3) to the target distribution π_β have been established in Wasserstein-2 distance under the assumptions that H is convex and (locally) Lipschitz continuous, see [3], [6], [12] and references therein. Recently, these results are considered under more relaxed conditions aiming to include a wider range of practical applications:

(i) To relax the convexity condition on H, one line of research is to replace such an assumption with a dissipativity condition. In [42], a convergence result is obtained in Wasserstein-2 distance between the law of the SGLD algorithm (1.3) and π_β with the rate $\gamma^{5/4}n$. This is the first such result in non-convex optimization, which is then improved in the work [52] and [9]. Compared to [42], a higher rate of convergence with dependence on n is achieved in [52] following a direct analysis of the ergodicity of the overdamped LMC algorithms; whereas a rate $1/2$ (independent of n) in Wasserstein-1 distance is obtained in [9] by using the contraction results developed in [17]. Another line of research is to replace the convexity condition with a convexity at infinity condition as discussed in [10] and [34], where the convergence results in Wasserstein-1 distance are obtained by using the contraction property in [16].

(ii) As for the relaxation of the smoothness of H, relevant convergence results have been established for the stochastic gradient descent (SGD) algorithm with discontinuous drift, see [18] and [8]. In particular, [18] provides an almost sure convergence result, while [8] provides a strong $\mathscr{L}^1$ convergence result with rate $1/2$.

In Chapter 4, we consider the SGLD algorithm (1.3) with discontinuous gradient H. By extending the techniques used in [3], [8] and [9], non-asymptotic estimates are obtained in Wasserstein distances between the law of the SGLD algorithm (1.3) and π_β in both convex and non-convex case. Examples are then provided to illustrate the wide applicability of the proposed algorithm in statistics, machine learning and finance.

1.2 Notation

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Fix an integer $d \geq 1$. For an $\mathbb{R}^d$-valued random variable Z, its law on $\mathscr{B}(\mathbb{R}^d)$ (the Borel sigma-algebra of $\mathbb{R}^d$) is denoted by $\mathcal{L}(Z)$, and we denote by $\mathbb{E}[Z]$ its expectation. For $1 \leq p < \infty$, denote by $\mathscr{L}^p$ the usual space of p-integrable real-valued random variables. The Euclidean norm of a vector $b \in \mathbb{R}^d$, the spectral norm and the Frobenius norm of a matrix $\sigma \in \mathbb{R}^{d\times m}$ are denoted by $|b|$, $|\sigma|$ and $|\sigma|_\mathsf{F}$ respectively. σ^T is the transpose matrix of σ. The i-th element of b and (i,j)-th element of σ are denoted respectively by $b^{(i)}$ and $\sigma^{(i,j)}$, for every $i = 1, \dots, d$ and $j = 1, \dots, m$. In addition, denote by $\lfloor a \rfloor$ the integer part of a positive real number a, and $\lceil a \rceil = \lfloor a \rfloor + 1$. The inner product of two vectors $x, y \in \mathbb{R}^d$ is denoted by $\langle x, y\rangle$. For all $x \in \mathbb{R}^d$ and $M > 0$, denote by $\mathrm{B}(x, M)$ (respectively $\overline{\mathrm{B}}(x, M)$) the open (respectively close) ball centered at x with radius M. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a twice continuously differentiable function. Denote by ∇f, $\nabla^2 f$ and Δf the gradient of f, the Hessian of f and the Laplacian of f respectively. Let $g : \mathbb{R}^d \to \mathbb{R}^d$ be a twice continuously differentiable function. Denote by $\vec{\Delta}g$ the vector Laplacian of g, i.e., for all $x \in \mathbb{R}^d$, $\vec{\Delta}g(x)$ is a vector in $\mathbb{R}^d$ whose i-th entry is $\sum_{u=1}^d \frac{\partial^2 g^{(i)}}{\partial x^{(u)} \partial x^{(u)}}(x)$. For $m \geq 1$, let $f : \mathbb{R}^d \to \mathbb{R}^m$ be a continuous function, and define $C_{\text{poly}}(\mathbb{R}^d, \mathbb{R}^m)$ the set of continuous functions such that for $f \in C_{\text{poly}}(\mathbb{R}^d, \mathbb{R}^m)$, $|f(x)| \leq C_q(1+|x|^q)$ for all $x \in \mathbb{R}^d$ with $C_q, q \geq 0$. For any $t \geq 0$, denote by $\mathcal{C}([0,t], \mathbb{R}^d)$ the space of continuous $\mathbb{R}^d$-valued paths defined on the time interval $[0,t]$.

For any integer $d \geq 1$, let $\mathscr{P}(\mathbb{R}^d)$ denote the set of probability measures on $\mathscr{B}(\mathbb{R}^d)$. For $\mu \in \mathscr{P}(\mathbb{R}^d)$ and for a non-negative measurable $f : \mathbb{R}^d \to \mathbb{R}$, we denote by $\mu(f) := \int_{\mathbb{R}^d} f(\theta)\mu(d\theta)$. Given a Markov kernel R on $\mathbb{R}^d$ and a function f integrable under $R(x, \cdot)$, denote by $Rf(x) = \int_{\mathbb{R}^d} f(y)R(x, dy)$. Let $V : \mathbb{R}^d \to [1, \infty)$ be a measurable function. The V-total variation distance between μ and ν is defined as $\|\mu - \nu\|_V = \sup_{|f| \leq V} |\mu(f) - \nu(f)|$. If $V = 1$, then $\|\cdot\|_V$ is the total variation denoted by $\|\cdot\|_{TV}$. Let μ and ν be two probability measures on a state space Ω with a given σ-algebra. If $\mu \ll \nu$, we denote by $d\mu/d\nu$ the Radon-Nikodym derivative of μ w.r.t. ν. Then, the Kullback-Leibler divergence of μ w.r.t. ν is given by

$$\mathrm{KL}(\mu|\nu) = \int_\Omega \frac{d\mu}{d\nu} \log\left(\frac{d\mu}{d\nu}\right) d\nu.$$

For $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$, let $\mathscr{C}(\mu, \nu)$ denote the set of probability measures ζ on $\mathscr{B}(\mathbb{R}^{2d})$ such that its respective marginals are μ, ν. Furthermore, we say that a couple of $\mathbb{R}^d$-valued random variables (X, Y) is a coupling of μ and ν if there exists $\zeta \in \mathscr{C}(\mu, \nu)$ such that (X, Y) is distributed according to ζ. For two probability measures μ and ν, the Wasserstein distance of order $p \geq 1$ is defined as

$$W_p(\mu, \nu) = \left(\inf_{\zeta \in \mathscr{C}(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \, d\zeta(x, y) \right)^{1/p}, \quad \mu, \nu \in \mathscr{P}(\mathbb{R}^d).$$

1.3 Useful inequalities

In this section, we present some inequalities that are frequently used in this thesis.

Lemma 1.1. *(Young's inequality) Let a, b be non-negative real numbers. Then, for $1 < p, q < \infty$ with $1/p + 1/q = 1$,*

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Lemma 1.2. *(Hölder's inequality) Let $(X, \mathcal{A}, \mu)$ be a measure space. Then, for all measurable functions $f, g : \mathbb{X} \to \mathbb{R}$, and for any $1 < p, q < \infty$ with $1/p + 1/q = 1$,*

$$\int_X |f(x)g(x)| d\mu(x) \leq \left(\int_X |f(x)|^p d\mu(x) \right)^{\frac{1}{p}} \left(\int_X |g(x)|^q d\mu(x) \right)^{\frac{1}{q}}.$$

Lemma 1.3. *(Gronwall's lemma) Let $T > 0$ and $C > 0$ be fixed constants. Let $u : [0, T] \to \mathbb{R}_+$ be a bounded Borel measurable function and $v : [0, T] \to \mathbb{R}_+$ be an integrable function. Then, if*

$$u(t) \leq C + \int_0^t v(s)u(s)ds < \infty,$$

for all $0 \leq t \leq T$, then

$$u(t) \leq c \exp\left(\int_0^t v(s)ds \right).$$

Lemma 1.4. *(Burkholder-Davis-Gundy inequality) Let $M := (M_t)_{t \geq 0}$ be a continuous local martingale with $M_0 = 0$ on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ equipped with the filtration $(\mathscr{F}_t)_{t \geq 0}$. Denote by $[M]$ the quadratic variation of M and $M_t^* := \sup_{0 \leq s \leq t} |M_s|$ its maximum process. Then, for any $0 < p < \infty$, there exist positive constants c_p and C_p such that*

$$c_p \mathbb{E}\left([M]_\tau\right)^{\frac{p}{2}} \leq \mathbb{E}\left(M_\tau^*\right)^p \leq C_p \mathbb{E}\left([M]_\tau\right)^{\frac{p}{2}}$$

for every stopping time τ.

Lemma 1.5. *(Jensen's inequality) Let* X *be an integrable real-valued random variable on a probability space* $(\Omega, \mathscr{F}, \mathbb{P})$ *and let* φ *be a convex function. Then,*

$$\varphi\left(\mathbb{E}[X]\right) \leq \mathbb{E}\left[\varphi(X)\right].$$

Lemma 1.6. *(Conditional Jensen's inequality) Let* X *be an integrable real-valued random variable on a probability space* $(\Omega, \mathscr{F}, \mathbb{P})$*. Let* φ *be a convex function and* $\mathscr{G} \subset \mathscr{F}$ *be a* σ*-algebra. Then, almost surely,*

$$\varphi\left(\mathbb{E}[X|\mathscr{G}]\right) \leq \mathbb{E}\left[\varphi(X)|\mathscr{G}\right].$$

Chapter 2

Tamed order 1.5 algorithm

2.1 Introduction

In this chapter, a new type of explicit order 1.5 scheme is constructed to approximate SDEs with super-linear coefficients. The main idea is to follow the approach of [41] by applying an appropriate Ito-Taylor (known also as Wagner-Platen) expansion and to use the taming technique introduced in [45] and [30]. Under certain conditions (given below), an $\mathscr{L}^2$ convergence result of the proposed order 1.5 scheme is obtained in Theorem 2.6 by extending the techniques used in [45] and [30].

This chapter is based on my joint work [47]. It is organised as follows. Section 2.2 presents the assumptions and main results. Section 2.3 provides the $\mathscr{L}^p$ moment bounds for the proposed order 1.5 algorithm. The proof of the main result is presented in Section 2.4, which is followed by numerical experiments in Section 2.5. Auxiliary results are provided in Appendix A.

2.2 Main results

Let $(\Omega, \{\mathscr{F}_t\}_{t\geq 0}, \mathscr{F}, \mathbb{P})$ be a complete filtered probability space satisfying the usual conditions, which means that the filtration is right continuous and $\mathscr{F}_0$ contains all $\mathbb{P}$-null sets. Denote by $(w_t)_{t\in[0,T]}$ an m-dimensional Wiener process. Moreover, assume that b and σ are Borel-measurable functions from $\mathbb{R}^d$ to $\mathbb{R}^d$ and $\mathbb{R}^{d\times m}$, respectively. The drift and diffusion coefficients b and σ are assumed to be twice continuously differentiable in $x \in \mathbb{R}^d$. For a fixed $T > 0$, consider a d-dimensional SDE,

$$x_t = x_0 + \int_0^t b(x_s)\, ds + \int_0^t \sigma(x_s)\, dw_s, \tag{2.1}$$

almost surely for any $t \in [0, T]$, where x_0 is an $\mathscr{F}_0$-measurable random variable. We further introduce the following notation, which are frequently used in this chapter. For every $j = 1, \ldots, m$, define $L^0 : C^2(\mathbb{R}^d) \to C(\mathbb{R}^d)$ and $L^j : C^2(\mathbb{R}^d) \to C^1(\mathbb{R}^d)$ by

$$L^0 = \sum_{u=1}^d b^{(u)} \frac{\partial}{\partial x^{(u)}} + \frac{1}{2} \sum_{u,l=1}^d \sum_{j_1=1}^m \sigma^{(u,j_1)} \sigma^{(l,j_1)} \frac{\partial^2}{\partial x^{(u)} \partial x^{(l)}}, \quad L^j = \sum_{u=1}^d \sigma^{(u,j)} \frac{\partial}{\partial x^{(u)}}.$$

Moreover, for every $j, j_1 = 1, \ldots, m$, define $L^j L^{j_1} : C^2(\mathbb{R}^d) \to C(\mathbb{R}^d)$ by

$$L^j L^{j_1} = \sum_{u,l=1}^d \sigma^{(u,j)} \frac{\partial}{\partial x^{(u)}} \sigma^{(l,j_1)} \frac{\partial}{\partial x^{(l)}} + \sum_{u,l=1}^d \sigma^{(u,j)} \sigma^{(l,j_1)} \frac{\partial^2}{\partial x^{(u)} \partial x^{(l)}}.$$

Let $p_0 \geq 4$, $p_1 > 2$, and $\rho \geq 2$. The following assumptions are stated.

A-1 $\mathbb{E}|x_0|^{p_0} < \infty$.

A-2 There exists a constant $K > 0$, such that for any $x \in \mathbb{R}^d$,

$$2\langle x, b(x)\rangle + (p_0 - 1)|\sigma(x)|_{\mathsf{F}}^2 \leq K(1 + |x|^2).$$

A-3 There exists a constant $K > 0$, such that for any $x, \bar{x} \in \mathbb{R}^d$,

$$2\langle x - \bar{x}, b(x) - b(\bar{x})\rangle + (p_1 - 1)|\sigma(x) - \sigma(\bar{x})|_{\mathsf{F}}^2 \leq K|x - \bar{x}|^2.$$

A-4 There exists a constant $K > 0$, such that for any $x, \bar{x} \in \mathbb{R}^d$, and $i = 1, \ldots, d$,

$$|\nabla^2 b^{(i)}(x) - \nabla^2 b^{(i)}(\bar{x})| \leq K(1 + |x| + |\bar{x}|)^{\rho - 2}|x - \bar{x}|.$$

A-5 There exist constants $K > 0$ and $\alpha \in (0, 1]$, such that for any $x, \bar{x} \in \mathbb{R}^d$, $i = 1, \ldots, d$, and $j = 1, \ldots, m$,

$$|\nabla^2 \sigma^{(i,j)}(x) - \nabla^2 \sigma^{(i,j)}(\bar{x})| \leq K(1 + |x| + |\bar{x}|)^{\frac{\rho - 4}{2}}|x - \bar{x}|^{\alpha}.$$

Remark 2.1. *For a given matrix $Q \in \mathbb{R}^{d \times m}$, we have $|Q| \leq |Q|_{\mathsf{F}} \leq \sqrt{\min\{d, m\}}|Q|$.*

Remark 2.2. *Throughout this chapter, the constant $C > 0$ may take different values at different places, but it is always independent of $n \in \mathbb{N}$.*

Remark 2.3. *Assume **A-4** and **A-5** hold. Then, one can obtain the following estimates in a straightforward manner. In particular, by **A-4**, there exists a constant $C > 0$, such that for any $i, u, l = 1, \ldots, d$, and $x, \bar{x} \in \mathbb{R}^d$,*

$$\left|\frac{\partial^2 b^{(i)}(x)}{\partial x^{(u)} \partial x^{(l)}}\right| \leq C(1 + |x|)^{\rho - 1}.$$

In addition, there exists a constant $C > 0$

$$\left|\frac{\partial b^{(i)}(x)}{\partial x^{(u)}} - \frac{\partial b^{(i)}(\bar{x})}{\partial x^{(u)}}\right| \leq C(1 + |x| + |\bar{x}|)^{\rho - 1}|x - \bar{x}|.$$

Furthermore, there is a constant $C > 0$ such that for any $i, u = 1, \ldots, d$, and $x, \bar{x} \in \mathbb{R}^d$,

$$\left|\frac{\partial b^{(i)}(x)}{\partial x^{(u)}}\right| \leq C(1 + |x|)^{\rho},$$

$$|b(x) - b(\bar{x})| \leq C(1 + |x| + |\bar{x}|)^{\rho}|x - \bar{x}|,$$

which implies

$$|b(x)| \leq C(1 + |x|)^{\rho + 1}.$$

*Similarly, by **A-5**, there exists $C > 0$, such that for any $i, u, l = 1, \ldots, d$, $j = 1, \ldots, m$ and $x \in \mathbb{R}^d$,*

$$\left|\frac{\partial^2 \sigma^{(i,j)}(x)}{\partial x^{(u)} \partial x^{(l)}}\right| \leq C(1 + |x|)^{\frac{\rho - 2}{2}},$$

Moreover, there exists $C > 0$, such that for any $j = 1, \ldots, m$ and $x, \bar{x} \in \mathbb{R}^d$,

$$\left|\frac{\partial \sigma^{(i,j)}(x)}{\partial x^{(u)}} - \frac{\partial \sigma^{(i,j)}(\bar{x})}{\partial x^{(u)}}\right| \leq C(1 + |x| + |\bar{x}|)^{\frac{\rho - 2}{2}}|x - \bar{x}|.$$

Furthermore, there exists $C > 0$, such that for any $i, u = 1, \ldots, d$, $j = 1, \ldots, m$ and $x, \bar{x} \in \mathbb{R}^d$,

$$\left|\frac{\partial \sigma^{(i,j)}(x)}{\partial x^{(u)}}\right| \leq C(1 + |x|)^{\frac{\rho}{2}},$$

$$|\sigma(x) - \sigma(\bar{x})| \leq C(1 + |x| + |\bar{x}|)^{\frac{\rho}{2}}|x - \bar{x}|,$$

which implies

$$|\sigma(x)| \le C(1+|x|)^{\frac{\rho}{2}+1}.$$

Then, there exists a constant $C>0$, such that

$$|L^0 b(x)| \le C(1+|x|)^{2\rho+1}, \quad |L^j b(x)| \le C(1+|x|)^{\frac{3}{2}\rho+1},$$

$$|L^0 \sigma(x)| \le C(1+|x|)^{\frac{3}{2}\rho+1}, \quad |L^j \sigma(x)| \le C(1+|x|)^{\rho+1},$$

$$|L^j L^{j_1} \sigma(x)| \le C(1+|x|)^{\frac{3}{2}\rho+1}.$$

We adopt a uniform taming approach meaning that all terms of interest in the numerical scheme, which are used to approximate the SDE (2.1), are controlled in the same way, i.e. $\frac{1}{1+n^{-\theta}|x|^{2\rho\theta}}$ is used where θ represents the desired rate. More concretely, in the order 1.5 paradigm, one constructs, for any $n \in \mathbb{N}$ and $f \in C^2(\mathbb{R}^d)$,

$$f^n(x) = \frac{f(x)}{1+n^{-\theta}|x|^{2\rho\theta}}, \quad L^{n,0} f(x) := \frac{L^0 f(x)}{1+n^{-\theta}|x|^{2\rho\theta}},$$

$$L^{n,j} f(x) := \frac{L^j f(x)}{1+n^{-\theta}|x|^{2\rho\theta}}, \quad L^{n,j} L^{j_1} f(x) := \frac{L^j L^{j_1} f(x)}{1+n^{-\theta}|x|^{2\rho\theta}},$$

where θ is taken to be $3/2$.

Remark 2.4. *Due to Remark 2.3, one observes that, there exists a constant $C>0$, such that for any $n \in \mathbb{N}$*

$$|b^n(x)| \le \min\left(Cn^{\frac{1}{2}}(1+|x|), |b(x)|\right), \quad |\sigma^n(x)|^2 \le \min\left(Cn^{\frac{1}{2}}(1+|x|^2), |\sigma(x)|^2\right),$$

$$|L^{n,0} b(x)| \le \min\left(Cn(1+|x|), |L^0 b(x)|\right), \quad |L^{n,j} b(x)| \le \min\left(Cn^{\frac{3}{4}}(1+|x|), |L^j b(x)|\right),$$

$$|L^{n,0} \sigma(x)| \le \min\left(Cn^{\frac{3}{4}}(1+|x|), |L^0 \sigma(x)|\right), \quad |L^{n,j} \sigma(x)| \le \min\left(Cn^{\frac{1}{2}}(1+|x|), |L^j \sigma(x)|\right),$$

$$|L^{n,j} L^{j_1} \sigma(x)| \le \min\left(Cn^{\frac{3}{4}}(1+|x|), |L^j L^{j_1} \sigma(x)|\right).$$

Define $\kappa(n,t) := \lfloor nt \rfloor / n$, for any $n \in \mathbb{N}$ and $t \in [0,T]$. Denote by

$$b_1^n(t,x) = \int_{\kappa(n,t)}^t L^{n,0} b(x)\, ds, \quad b_2^n(t,x) = \sum_j \int_{\kappa(n,t)}^t L^{n,j} b(x)\, dw_s^j,$$

$$\tilde{b}^n(t,x) = b^n(x) + b_1^n(t,x) + b_2^n(t,x),$$

$$\sigma_1^n(t,x) = \sum_j \int_{\kappa(n,t)}^t L^{n,j} \sigma(x)\, dw_s^j, \quad \sigma_2^n(t,x) = \int_{\kappa(n,t)}^t L^{n,0} \sigma(x)\, ds,$$

$$\sigma_3^n(t,x) = \sum_j \sum_{j_1} \int_{\kappa(n,t)}^t \int_{\kappa(n,t)}^s L^{n,j} L^{j_1} \sigma(x)\, dw_r^j\, dw_s^{j_1},$$

$$\tilde{\sigma}^n(t,x) = \sigma^n(x) + \sigma_M^n(t,x),$$

where $\sigma_M^n(t,x) = \sigma_1^n(t,x) + \sigma_2^n(t,x) + \sigma_3^n(t,x)$. The tamed order 1.5 strong Taylor scheme is as follows:

$$x_t^n = x_0 + \int_0^t \tilde{b}^n(s, x_{\kappa(n,s)}^n)\, ds + \int_0^t \tilde{\sigma}^n(s, x_{\kappa(n,s)}^n)\, dw_s, \tag{2.2}$$

almost surely for any $t \in [0,T]$.

Remark 2.5. *For the numerical implementation, one may consider using the discrete version of the tamed order 1.5 scheme (2.2), which is given explicitly in Section 2.5 with $d=m=1$. As for the high dimensional case, the diffusion coefficient needs to satisfy the commutative condition, see e.g. [27, Chapter 10.4], and the corresponding discrete version of the proposed scheme (2.2) is given in [27, Chapter 10.4 (4.15)] with the coefficients $b(x), \sigma(x)$ replaced by*

$b^n(x), \sigma^n(x)$. *Otherwise, one needs to handle the associated Levy areas. One possible approach is to use a coupling technique (see [13]).*

Theorem 2.6. *Assume **A-1** - **A-5** are satisfied with $p_0 \geq 2(5\rho+1)$, then the explicit order 1.5 scheme* (2.2) *converges to the true solution of the SDE* (2.1) *in $\mathscr{L}^2$ with a rate of convergence equal to $1+\alpha/2$, i.e., there exists a constant $C>0$, such that for any $n \in \mathbb{N}$,*

$$\left(\sup_{0\leq t\leq T} \mathbb{E}|x_t - x_t^n|^2\right)^{1/2} \leq C n^{-(1+\alpha/2)}. \tag{2.3}$$

2.3 Moment bounds

Lemma 2.7. *Assume **A-1** - **A-3** hold. Then, there is a unique solution to the SDE* (2.1)*, and the p_0-th moment of the solution is bounded uniformly in time, i.e. there exists a constant $C>0$, such that for any $t \in [0,T]$,*

$$\sup_{0\leq t\leq T} \mathbb{E}|x_t|^{p_0} \leq C.$$

Proof. It is a well-known result, and the proof can be found in [35]. □

Remark 2.8. *By Remark 2.4, for each $n \in \mathbb{N}$, the norm of $\tilde{b}^n$ and $\tilde{\sigma}^n$ are growing at most linearly in x. Then, together with **A-1**, this guarantees that for any $n \in \mathbb{N}$ and $p \leq p_0$,*

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |x_t^n|^p\right] < \infty.$$

Lemma 2.9. *Let **A-4** - **A-5** be satisfied, then there exists a constant $C>0$, such that for any $n \in \mathbb{N}$ and $t \in [0,T]$,*

$$\mathbb{E}|b_1^n(t, x_{\kappa(n,t)}^n)|^{p_0} \leq C(1+\mathbb{E}|x_{\kappa(n,t)}^n|^{p_0}),$$
$$\mathbb{E}|b_2^n(t, x_{\kappa(n,t)}^n)|^{p_0} \leq C n^{\frac{p_0}{4}}(1+\mathbb{E}|x_{\kappa(n,t)}^n|^{p_0}),$$
$$\mathbb{E}|\sigma_1^n(t, x_{\kappa(n,t)}^n)|^{p_0} \leq C(1+\mathbb{E}|x_{\kappa(n,t)}^n|^{p_0}),$$
$$\mathbb{E}|\sigma_2^n(t, x_{\kappa(n,t)}^n)|^{p_0} \leq C(1+\mathbb{E}|x_{\kappa(n,t)}^n|^{p_0}),$$
$$\mathbb{E}|\sigma_3^n(t, x_{\kappa(n,t)}^n)|^{p_0} \leq C(1+\mathbb{E}|x_{\kappa(n,t)}^n|^{p_0}).$$

Proof. Due to Remark 2.4, these inequalities follow immediately. □

Corollary 2.10. *Assume **A-4** - **A-5** are satisfied, then there exists a constant $C>0$, such that for any $n \in \mathbb{N}$ and $t \in [0,T]$,*

$$\mathbb{E}|\tilde{b}^n(t, x_{\kappa(n,t)}^n)|^{p_0} \leq C n^{\frac{p_0}{2}}(1+\mathbb{E}|x_{\kappa(n,t)}^n|^{p_0}),$$

$$\mathbb{E}|\tilde{\sigma}^n(t, x_{\kappa(n,t)}^n)|^{p_0} \leq C n^{\frac{p_0}{4}}(1+\mathbb{E}|x_{\kappa(n,t)}^n|^{p_0}).$$

Lemma 2.11. *Assume **A-1** - **A-5** hold, then there exists a constant $C>0$, such that for any $n \in \mathbb{N}$, the order* 1.5 *scheme* (2.2) *satisfies*

$$\sup_{n\in\mathbb{N}} \sup_{0\leq t\leq T} \mathbb{E}|x_t^n|^{p_0} \leq C.$$

Proof. Itô's formula gives, almost surely,

$$\begin{aligned}|x_t^n|^{p_0} &= |x_0|^{p_0} + p_0 \int_0^t |x_s^n|^{p_0-2} \left\langle x_s^n, \tilde{b}^n(s, x_{\kappa(n,s)}^n)\right\rangle ds \\ &\quad + p_0 \int_0^t |x_s^n|^{p_0-2} \left\langle x_s^n, \tilde{\sigma}^n(s, x_{\kappa(n,s)}^n)\, dw_s\right\rangle\end{aligned}$$

$$+\frac{p_0}{2}\int_0^t |x_s^n|^{p_0-2}|\tilde{\sigma}^n(s,x^n_{\kappa(n,s)})|_{\mathsf{F}}^2\,ds$$
$$+\frac{p_0(p_0-2)}{2}\int_0^t |x_s^n|^{p_0-4}|(\tilde{\sigma}^n(s,x^n_{\kappa(n,s)}))^{\mathsf{T}}x_s^n|^2\,ds,$$

for any $t\in[0,T]$. Then, since the expectation of the third term above is zero, one obtains

$$\begin{aligned}\mathbb{E}|x_t^n|^{p_0}\leq&\mathbb{E}|x_0|^{p_0}+p_0\mathbb{E}\int_0^t|x_s^n|^{p_0-2}\left\langle(x_s^n-x^n_{\kappa(n,s)}),b^n(x^n_{\kappa(n,s)})\right\rangle ds\\&+p_0\mathbb{E}\int_0^t|x_s^n|^{p_0-2}\left\langle x^n_{\kappa(n,s)},b^n(x^n_{\kappa(n,s)})\right\rangle ds\\&+p_0\mathbb{E}\int_0^t|x_s^n|^{p_0-2}\left\langle x^n_s,b_1^n(s,x^n_{\kappa(n,s)})\right\rangle ds\\&+p_0\mathbb{E}\int_0^t|x_s^n|^{p_0-2}\left\langle x^n_s,b_2^n(s,x^n_{\kappa(n,s)})\right\rangle ds\\&+\frac{p_0(p_0-1)}{2}\mathbb{E}\int_0^t|x_s^n|^{p_0-2}|\tilde{\sigma}^n(s,x^n_{\kappa(n,s)})|_{\mathsf{F}}^2\,ds,\end{aligned}$$

which can be written as

$$\mathbb{E}|x_t^n|^{p_0}\leq G_1+\sum_{i=2}^7 G_i(t),\tag{2.4}$$

where $G_1=\mathbb{E}|x_0|^{p_0}$,

$$\begin{aligned}G_2(t)&=p_0\mathbb{E}\int_0^t|x_s^n|^{p_0-2}\left\langle(x_s^n-x^n_{\kappa(n,s)}),b^n(x^n_{\kappa(n,s)})\right\rangle ds,\\G_3(t)&=\frac{p_0}{2}\mathbb{E}\int_0^t|x_s^n|^{p_0-2}(2\left\langle x^n_{\kappa(n,s)},b^n(x^n_{\kappa(n,s)})\right\rangle+(p_0-1)|\sigma^n(x^n_{\kappa(n,s)})|_{\mathsf{F}}^2)\,ds,\\G_4(t)&=p_0\mathbb{E}\int_0^t|x_s^n|^{p_0-2}\left\langle x^n_s,b_1^n(s,x^n_{\kappa(n,s)})\right\rangle ds,\\G_5(t)&=p_0\mathbb{E}\int_0^t|x_s^n|^{p_0-2}\left\langle x^n_s,b_2^n(s,x^n_{\kappa(n,s)})\right\rangle ds,\\G_6(t)&=\frac{p_0(p_0-1)}{2}\mathbb{E}\int_0^t|x_s^n|^{p_0-2}|\sigma_M^n(s,x^n_{\kappa(n,s)})|_{\mathsf{F}}^2\,ds,\\G_7(t)&=p_0(p_0-1)\mathbb{E}\int_0^t|x_s^n|^{p_0-2}\sum_{k=1}^d\sum_{v=1}^m\sigma^{n,(k,v)}(x^n_{\kappa(n,s)})\sigma_M^{n,(k,v)}(s,x^n_{\kappa(n,s)})\,ds.\end{aligned}$$

In order to estimate $G_2(t)$, one writes

$$\begin{aligned}G_2(t)=&p_0\mathbb{E}\int_0^t|x_s^n|^{p_0-2}\left\langle\int_{\kappa(n,s)}^s\tilde{b}^n(r,x^n_{\kappa(n,r)})\,dr,b^n(x^n_{\kappa(n,s)})\right\rangle ds\\&+p_0\mathbb{E}\int_0^t|x_s^n|^{p_0-2}\left\langle\int_{\kappa(n,s)}^s\tilde{\sigma}^n(r,x^n_{\kappa(n,r)})\,dw_r,b^n(x^n_{\kappa(n,s)})\right\rangle ds,\end{aligned}$$

for any $t\in[0,T]$. By applying Young's inequality and Remark 2.4, the following estimate can be obtained

$$G_2(t)\leq C+C\int_0^t\sup_{0\leq r\leq s}\mathbb{E}|x_r^n|^{p_0}ds+C\mathbb{E}\int_0^t\left|n^{\frac{1}{2}}\int_{\kappa(n,s)}^s\tilde{b}^n(r,x^n_{\kappa(n,r)})\,dr\right|^{p_0}ds$$

$$
\begin{aligned}
&+ p_0\mathbb{E}\int_0^t (|x_s^n|^{p_0-2} - |x_{\kappa(n,s)}^n|^{p_0-2}|) \left\langle \int_{\kappa(n,s)}^s \tilde{\sigma}^n(r, x_{\kappa(n,r)}^n)\, dw_r, b^n(x_{\kappa(n,s)}^n) \right\rangle ds \\
&+ p_0\mathbb{E}\int_0^t |x_{\kappa(n,s)}^n|^{p_0-2} \left\langle \int_{\kappa(n,s)}^s \tilde{\sigma}^n(r, x_{\kappa(n,r)}^n)\, dw_r, b^n(x_{\kappa(n,s)}^n) \right\rangle ds,
\end{aligned}
$$

for any $t \in [0,T]$. Since the last term above is zero, by using Corollary 2.10, Itô's formula and by the fact that for any matrix $A \in \mathbb{R}^{d\times m}$, $|A|_{\mathsf{F}} \leq C|A|$ for some $C > 0$, it follows that,

$$
\begin{aligned}
G_2(t) \leq\ & C + C\int_0^t \sup_{0\leq r\leq s} \mathbb{E}|x_r^n|^{p_0} ds \\
&+ C\mathbb{E}\int_0^t \int_{\kappa(n,s)}^s |x_r^n|^{p_0-4} \left\langle x_r^n, \tilde{b}^n(r, x_{\kappa(n,r)}^n) \right\rangle dr \\
&\qquad\qquad \times \left\langle \int_{\kappa(n,s)}^s \tilde{\sigma}^n(r, x_{\kappa(n,r)}^n)\, dw_r, b^n(x_{\kappa(n,s)}^n) \right\rangle ds \\
&+ C\mathbb{E}\int_0^t \int_{\kappa(n,s)}^s |x_r^n|^{p_0-4} \left\langle x_r^n, \tilde{\sigma}^n(r, x_{\kappa(n,r)}^n)\, dw_r \right\rangle \\
&\qquad\qquad \times \left\langle \int_{\kappa(n,s)}^s \tilde{\sigma}^n(r, x_{\kappa(n,r)}^n)\, dw_r, b^n(x_{\kappa(n,s)}^n) \right\rangle ds \\
&+ C\mathbb{E}\int_0^t \int_{\kappa(n,s)}^s |x_r^n|^{p_0-4} |\tilde{\sigma}^n(r, x_{\kappa(n,r)}^n)|^2\, dr \\
&\qquad\qquad \times \left| \int_{\kappa(n,s)}^s \tilde{\sigma}^n(r, x_{\kappa(n,r)}^n)\, dw_r \right| |b^n(x_{\kappa(n,s)}^n)|\, ds.
\end{aligned}
$$

Due to Remark 2.4,

$$
\begin{aligned}
G_2(t) \leq\ & C + C\int_0^t \sup_{0\leq r\leq s} \mathbb{E}|x_r^n|^{p_0} ds \\
&+ Cn^{\frac{1}{2}}\mathbb{E}\int_0^t \int_{\kappa(n,s)}^s |x_r^n|^{p_0-3}(1+|x_{\kappa(n,s)}^n|)|\tilde{b}^n(r, x_{\kappa(n,r)}^n)|\, dr \left| \int_{\kappa(n,s)}^s \tilde{\sigma}^n(r, x_{\kappa(n,r)}^n)\, dw_r \right| ds \\
&+ Cn^{\frac{1}{2}}\mathbb{E}\int_0^t \int_{\kappa(n,s)}^s |x_r^n|^{p_0-3}(1+|x_{\kappa(n,s)}^n|)|\tilde{\sigma}^n(r, x_{\kappa(n,r)}^n)|^2\, dr\, ds \\
&+ Cn^{\frac{1}{2}}\mathbb{E}\int_0^t \int_{\kappa(n,s)}^s |x_r^n|^{p_0-4}(1+|x_{\kappa(n,s)}^n|)|\tilde{\sigma}^n(r, x_{\kappa(n,r)}^n)|^2\, dr \left| \int_{\kappa(n,s)}^s \tilde{\sigma}^n(r, x_{\kappa(n,r)}^n)\, dw_r \right| ds,
\end{aligned}
$$

for any $t \in [0,T]$. Then, the application of Young's inequality yields

$$
\begin{aligned}
G_2(t) \leq\ & C + C\int_0^t \sup_{0\leq r\leq s} \mathbb{E}|x_r^n|^{p_0} ds \\
&+ C\mathbb{E}\int_0^t n^{\frac{1}{4}} \int_{\kappa(n,s)}^s (1+|x_r^n|^{p_0-2}+|x_{\kappa(n,s)}^n|^{p_0-2})|\tilde{b}^n(r, x_{\kappa(n,r)}^n)|\, dr \\
&\qquad\qquad \times n^{\frac{1}{4}} \left| \int_{\kappa(n,s)}^s \tilde{\sigma}^n(r, x_{\kappa(n,r)}^n)\, dw_r \right| ds \\
&+ C\mathbb{E}\int_0^t \int_{\kappa(n,s)}^s n^{1-\frac{2}{p_0}}(1+|x_r^n|^{p_0-2}+|x_{\kappa(n,s)}^n|^{p_0-2}) n^{-\frac{1}{2}+\frac{2}{p_0}} |\tilde{\sigma}^n(r, x_{\kappa(n,r)}^n)|^2\, dr\, ds \\
&+ C\mathbb{E}\int_0^t n^{\frac{1}{4}} \int_{\kappa(n,s)}^s (1+|x_r^n|^{p_0-3}+|x_{\kappa(n,s)}^n|^{p_0-3})|\tilde{\sigma}^n(r, x_{\kappa(n,r)}^n)|^2\, dr
\end{aligned}
$$

$$\times n^{\frac{1}{4}} \left| \int_{\kappa(n,s)}^{s} \tilde{\sigma}^n(r, x^n_{\kappa(n,r)})\, dw_r \right| ds,$$

which can be further estimated as

$$\begin{aligned}
G_2(t) \le C &+ C\int_0^t \sup_{0\le r\le s} \mathbb{E}|x^n_r|^{p_0} ds \\
&+ C\mathbb{E}\int_0^t \left(\int_{\kappa(n,s)}^{s} n^{\frac{3}{4}-\frac{1}{p_0}} (1+|x^n_r|^{p_0-2} + |x^n_{\kappa(n,s)}|^{p_0-2}) n^{-\frac{1}{2}+\frac{1}{p_0}} |\tilde{b}^n(r, x^n_{\kappa(n,r)})|\, dr \right)^{\frac{p_0}{p_0-1}} ds \\
&+ Cn\mathbb{E}\int_0^t \int_{\kappa(n,s)}^{s} (1+|x^n_r|^{p_0} + |x^n_{\kappa(n,s)}|^{p_0})\, dr\, ds \\
&+ Cn^{-\frac{p_0}{4}+1}\mathbb{E}\int_0^t \int_{\kappa(n,s)}^{s} |\tilde{\sigma}^n(r, x^n_{\kappa(n,r)})|^{p_0}\, dr\, ds \\
&+ C\mathbb{E}\int_0^t \left(\int_{\kappa(n,s)}^{s} n^{\frac{3}{4}-\frac{2}{p_0}} (1+|x^n_r|^{p_0-3} + |x^n_{\kappa(n,s)}|^{p_0-3}) n^{-\frac{1}{2}+\frac{2}{p_0}} |\tilde{\sigma}^n(r, x^n_{\kappa(n,r)})|^2\, dr \right)^{\frac{p_0}{p_0-1}} ds \\
&+ Cn^{\frac{p_0}{4}} \int_0^t \mathbb{E}\left| \int_{\kappa(n,s)}^{s} \tilde{\sigma}^n(r, x^n_{\kappa(n,r)})\, dw_r \right|^{p_0} ds
\end{aligned}$$

for any $t \in [0, T]$. By using Young's inequality and Corollary 2.10, one obtains

$$\begin{aligned}
G_2(t) \le C &+ C\int_0^t \sup_{0\le r\le s} \mathbb{E}|x^n_r|^{p_0} ds \\
&+ C\mathbb{E}\int_0^t \left(\int_{\kappa(n,s)}^{s} n^{\frac{3p_0-4}{4p_0}\times\frac{p_0-1}{p_0-2}} (1+|x^n_r|^{p_0-1} + |x^n_{\kappa(n,s)}|^{p_0-1})\, dr \right)^{\frac{p_0}{p_0-1}} ds \\
&+ C\mathbb{E}\int_0^t \left(\int_{\kappa(n,s)}^{s} n^{\frac{(2-p_0)\times(p_0-1)}{2p_0}} |\tilde{b}^n(r, x^n_{\kappa(n,r)})|^{p_0-1}\, dr \right)^{\frac{p_0}{p_0-1}} ds \\
&+ C\mathbb{E}\int_0^t \left(\int_{\kappa(n,s)}^{s} n^{\frac{3p_0-8}{4p_0}\times\frac{p_0-1}{p_0-3}} (1+|x^n_r|^{p_0-1} + |x^n_{\kappa(n,s)}|^{p_0-1})\, dr \right)^{\frac{p_0}{p_0-1}} ds \\
&+ C\mathbb{E}\int_0^t \left(\int_{\kappa(n,s)}^{s} n^{\frac{4-p_0}{2p_0}\times\frac{p_0-1}{2}} |\tilde{\sigma}^n(r, x^n_{\kappa(n,r)})|^{p_0-1}\, dr \right)^{\frac{p_0}{p_0-1}} ds \\
&+ Cn^{-\frac{p_0}{4}+1} \int_0^t \int_{\kappa(n,s)}^{s} \mathbb{E}|\tilde{\sigma}^n(r, x^n_{\kappa(n,r)})|^{p_0}\, dr\, ds,
\end{aligned}$$

which, due to Hölder's inequality and Corollary 2.10, implies

$$\begin{aligned}
G_2(t) \le C &+ C\int_0^t \sup_{0\le r\le s} \mathbb{E}|x^n_r|^{p_0} ds \\
&+ Cn^{\frac{3p_0-4}{4(p_0-2)}-\frac{1}{p_0-1}} \int_0^t \mathbb{E}\int_{\kappa(n,s)}^{s} (1+|x^n_r|^{p_0} + |x^n_{\kappa(n,s)}|^{p_0})\, dr\, ds \\
&+ Cn^{-\frac{p_0}{2}+1-\frac{1}{p_0-1}} \int_0^t \mathbb{E}\int_{\kappa(n,s)}^{s} |\tilde{b}^n(r, x^n_{\kappa(n,r)})|^{p_0}\, dr\, ds \\
&+ Cn^{\frac{3p_0-8}{4(p_0-3)}-\frac{1}{p_0-1}} \int_0^t \mathbb{E}\int_{\kappa(n,s)}^{s} (1+|x^n_r|^{p_0} + |x^n_{\kappa(n,s)}|^{p_0})\, dr\, ds,
\end{aligned}$$

for any $t \in [0, T]$. Note that in the third and fifth term above, $n^{\frac{3p_0-4}{4(p_0-2)}}$ and $n^{\frac{3p_0-8}{4(p_0-3)}}$ are less than n for all $p_0 \geq 4$. Thus, in view of Corollary 2.10, one obtains

$$G_2(t) \leq C + C \int_0^t \sup_{0 \leq r \leq s} \mathbb{E}|x_r^n|^{p_0} ds,$$

for any $t \in [0, T]$. For $G_3(t)$, applying **A-2** gives

$$\begin{aligned} G_3(t) &= \frac{p_0}{2} \mathbb{E} \int_0^t |x_s^n|^{p_0 - 2} \frac{2 \left\langle x^n_{\kappa(n,s)}, b(x^n_{\kappa(n,s)}) \right\rangle + (p_0 - 1)|\sigma(x^n_{\kappa(n,s)})|^2_{\mathsf{F}}}{1 + n^{-3/2}|x^n_{\kappa(n,s)}|^{3\rho}} ds \\ &\leq C\mathbb{E} \int_0^t |x_s^n|^{p_0-2}(1 + |x^n_{\kappa(n,s)}|^2) \, ds, \end{aligned}$$

which, due to Young's inequality, results in

$$G_3(t) \leq C + C \int_0^t \sup_{0 \leq r \leq s} \mathbb{E}|x_r^n|^{p_0} ds,$$

for any $t \in [0, T]$. To estimate $G_4(t)$, one uses Young's inequality to obtain

$$G_4(t) \leq C\mathbb{E} \int_0^t |x_s^n|^{p_0} \, ds + C\mathbb{E} \int_0^t |b_1^n(s, x^n_{\kappa(n,s)})|^{p_0} \, ds,$$

which implies due to Lemma 2.9,

$$G_4(t) \leq C + C \int_0^t \sup_{0 \leq r \leq s} \mathbb{E}|x_r^n|^{p_0} ds,$$

for any $t \in [0, T]$. Moreover, one writes

$$G_5(t) = \sum_{i=1}^{3} G_{5i}(t),$$

where

$$\begin{aligned} G_{51}(t) &= p_0 \mathbb{E} \int_0^t |x_s^n|^{p_0-2} \left\langle (x_s^n - x^n_{\kappa(n,s)}), b_2^n(s, x^n_{\kappa(n,s)}) \right\rangle ds, \\ G_{52}(t) &= p_0 \mathbb{E} \int_0^t (|x_s^n|^{p_0-2} - |x^n_{\kappa(n,s)}|^{p_0-2}) \left\langle x^n_{\kappa(n,s)}, b_2^n(s, x^n_{\kappa(n,s)}) \right\rangle ds, \\ G_{53}(t) &= p_0 \mathbb{E} \int_0^t |x^n_{\kappa(n,s)}|^{p_0-2} \left\langle x^n_{\kappa(n,s)}, b_2^n(s, x^n_{\kappa(n,s)}) \right\rangle ds. \end{aligned}$$

One then calculates the following

$$\begin{aligned} G_{51}(t) = &p_0 \mathbb{E} \int_0^t |x_s^n|^{p_0-2} \left\langle \int_{\kappa(n,s)}^s \tilde{b}^n(r, x^n_{\kappa(n,r)}) \, dr, b_2^n(s, x^n_{\kappa(n,s)}) \right\rangle ds \\ &+ p_0 \mathbb{E} \int_0^t |x_s^n|^{p_0-2} \left\langle \int_{\kappa(n,s)}^s \tilde{\sigma}^n(r, x^n_{\kappa(n,r)}) \, dw_r, b_2^n(s, x^n_{\kappa(n,s)}) \right\rangle ds, \end{aligned}$$

which implies, due to Young's inequality,

$$\begin{aligned} G_{51}(t) \leq &C\mathbb{E} \int_0^t |x_s^n|^{p_0} \, ds + C\mathbb{E} \int_0^t \left| \left\langle n^{\frac{1}{4}} \int_{\kappa(n,s)}^s \tilde{b}^n(r, x^n_{\kappa(n,r)}) \, dr, n^{-\frac{1}{4}} b_2^n(s, x^n_{\kappa(n,s)}) \right\rangle \right|^{\frac{p_0}{2}} ds \\ &+ C\mathbb{E} \int_0^t \left| \left\langle n^{\frac{1}{4}} \int_{\kappa(n,s)}^s \tilde{\sigma}^n(r, x^n_{\kappa(n,r)}) \, dw_r, n^{-\frac{1}{4}} b_2^n(s, x^n_{\kappa(n,s)}) \right\rangle \right|^{\frac{p_0}{2}} ds, \end{aligned}$$

for any $t \in [0,T]$. Then, on applying Young's inequality again, one obtains

$$\begin{aligned}G_{51}(t) \leq & C\mathbb{E}\int_0^t |x_s^n|^{p_0}\,ds + Cn^{\frac{p_0}{4}}\mathbb{E}\int_0^t \left|\int_{\kappa(n,s)}^s \tilde{b}^n(r, x^n_{\kappa(n,r)})\,dr\right|^{p_0} ds \\ & + Cn^{\frac{p_0}{4}}\mathbb{E}\int_0^t \left|\int_{\kappa(n,s)}^s \tilde{\sigma}^n(r, x^n_{\kappa(n,r)})\,dw_r\right|^{p_0} ds + Cn^{-\frac{p_0}{4}}\mathbb{E}\int_0^t |b_2^n(s, x^n_{\kappa(n,s)})|^{p_0}\,ds,\end{aligned}$$

which by using Hölder's inequality and Lemma 2.9 yields

$$\begin{aligned}G_{51}(t) \leq & C + C\int_0^t \sup_{0\leq r\leq s}\mathbb{E}|x_r^n|^{p_0} ds \\ & + Cn^{\frac{p_0}{4}-p_0+1}\int_0^t\int_{\kappa(n,s)}^s \mathbb{E}|\tilde{b}^n(r, x^n_{\kappa(n,r)})|^{p_0}\,dr\,ds \\ & + Cn^{\frac{p_0}{4}-\frac{p_0}{2}+1}\int_0^t\int_{\kappa(n,s)}^s \mathbb{E}|\tilde{\sigma}^n(r, x^n_{\kappa(n,r)})|^{p_0}\,dr\,ds,\end{aligned}$$

for any $t \in [0,T]$. Due to Corollary 2.10, one concludes that

$$G_{51}(t) \leq C + C\int_0^t \sup_{0\leq r\leq s}\mathbb{E}|x_r^n|^{p_0} ds, \tag{2.5}$$

for any $t \in [0,T]$. As for $G_{52}(t)$, by using Itô's formula, one obtains

$$\begin{aligned}G_{52}(t) \leq & C\mathbb{E}\int_0^t\int_{\kappa(n,s)}^s |x_r^n|^{p_0-4}\left\langle x_r^n, \tilde{b}^n(r, x^n_{\kappa(n,r)})\right\rangle dr \left\langle x^n_{\kappa(n,s)}, b_2^n(s, x^n_{\kappa(n,s)})\right\rangle ds \\ & + C\mathbb{E}\int_0^t\int_{\kappa(n,s)}^s |x_r^n|^{p_0-4}\left\langle x_r^n, \tilde{\sigma}^n(r, x^n_{\kappa(n,r)})\,dw_r\right\rangle \\ & \qquad \times\left\langle x^n_{\kappa(n,s)}, \sum_j\int_{\kappa(n,s)}^s L^{n,j}b(x^n_{\kappa(n,r)})\,dw_r^j\right\rangle ds \\ & + C\mathbb{E}\int_0^t\int_{\kappa(n,s)}^s |x_r^n|^{p_0-4}|\tilde{\sigma}^n(r, x^n_{\kappa(n,r)})|^2\,dr|x^n_{\kappa(n,s)}||b_2^n(s, x^n_{\kappa(n,s)})|\,ds,\end{aligned}$$

which, by Young's inequality, can be expressed as

$$\begin{aligned}G_{52}(t) \leq & C\int_0^t\mathbb{E}\int_{\kappa(n,s)}^s n^{\frac{3}{4}-\frac{1}{p_0}}(1+|x_r^n|^{p_0-2}+|x^n_{\kappa(n,s)}|^{p_0-2})n^{-\frac{1}{2}+\frac{1}{p_0}}|\tilde{b}^n(r, x^n_{\kappa(n,r)})|\,dr \\ & \qquad \times n^{-\frac{1}{4}}|b_2^n(s, x^n_{\kappa(n,s)})|\,ds \\ & + C\sum_{j=1}^m\int_0^t\mathbb{E}\int_{\kappa(n,s)}^s (1+|x_r^n|^{p_0-2}+|x^n_{\kappa(n,s)}|^{p_0-2})|\tilde{\sigma}^n(r, x^n_{\kappa(n,r)})||L^{n,j}b(x^n_{\kappa(n,r)})|\,dr\,ds \\ & + C\int_0^t\mathbb{E}\int_{\kappa(n,s)}^s n^{\frac{3}{4}-\frac{2}{p_0}}(1+|x_r^n|^{p_0-3}+|x^n_{\kappa(n,s)}|^{p_0-3})n^{-\frac{1}{2}+\frac{2}{p_0}}|\tilde{\sigma}^n(r, x^n_{\kappa(n,r)})|^2\,dr \\ & \qquad \times n^{-\frac{1}{4}}|b_2^n(s, x^n_{\kappa(n,s)})|\,ds,\end{aligned}$$

for any $t \in [0,T]$. One uses Young's inequality again and Remark 2.4 to obtain

$$\begin{aligned}& G_{52}(t) \\ & \leq C\int_0^t\mathbb{E}\left(\int_{\kappa(n,s)}^s n^{\frac{3}{4}-\frac{1}{p_0}}(1+|x_r^n|^{p_0-2}+|x^n_{\kappa(n,s)}|^{p_0-2})n^{-\frac{1}{2}+\frac{1}{p_0}}|\tilde{b}^n(r, x^n_{\kappa(n,r)})|\,dr\right)^{\frac{p_0}{p_0-1}} ds\end{aligned}$$

$$+ C\int_0^t \mathbb{E}\int_{\kappa(n,s)}^s n^{1-\frac{1}{p_0}}(1+|x_r^n|^{p_0-1}+|x_{\kappa(n,s)}^n|^{p_0-1})n^{-\frac{1}{4}+\frac{1}{p_0}}|\tilde{\sigma}^n(r,x_{\kappa(n,r)}^n)|\,dr\,ds$$

$$+ C\int_0^t \mathbb{E}\left(\int_{\kappa(n,s)}^s n^{\frac{3}{4}-\frac{2}{p_0}}(1+|x_r^n|^{p_0-3}+|x_{\kappa(n,s)}^n|^{p_0-3})n^{-\frac{1}{2}+\frac{2}{p_0}}|\tilde{\sigma}^n(r,x_{\kappa(n,r)}^n)|^2\,dr\right)^{\frac{p_0}{p_0-1}} ds$$

$$+ Cn^{-\frac{p_0}{4}}\int_0^t \mathbb{E}|b_2^n(s,x_{\kappa(n,s)}^n)|^{p_0}\,ds,$$

which implies due to Lemma 2.9

$$\begin{aligned}
G_{52}(t) \le\, & C\int_0^t \mathbb{E}\left(\int_{\kappa(n,s)}^s n^{\frac{3p_0-4}{4p_0}\times\frac{p_0-1}{p_0-2}}(1+|x_r^n|^{p_0-1}+|x_{\kappa(n,s)}^n|^{p_0-1})\,dr\right)^{\frac{p_0}{p_0-1}} ds\\
&+ C\int_0^t \mathbb{E}\left(\int_{\kappa(n,s)}^s n^{\frac{(2-p_0)\times(p_0-1)}{2p_0}}|\tilde{b}^n(r,x_{\kappa(n,r)}^n)|^{p_0-1}\,dr\right)^{\frac{p_0}{p_0-1}} ds\\
&+ C\int_0^t \mathbb{E}\int_{\kappa(n,s)}^s n(1+|x_r^n|^{p_0}+|x_{\kappa(n,s)}^n|^{p_0})\,dr\,ds\\
&+ C\int_0^t \mathbb{E}\int_{\kappa(n,s)}^s n^{\frac{4-p_0}{4p_0}\times p_0}|\tilde{\sigma}^n(r,x_{\kappa(n,r)}^n)|^{p_0}\,dr\,ds\\
&+ C\int_0^t \mathbb{E}\left(\int_{\kappa(n,s)}^s n^{\frac{3p_0-8}{4p_0}\times\frac{p_0-1}{p_0-3}}(1+|x_r^n|^{p_0-1}+|x_{\kappa(n,s)}^n|^{p_0-1})\,dr\right)^{\frac{p_0}{p_0-1}} ds\\
&+ C\int_0^t \mathbb{E}\left(\int_{\kappa(n,s)}^s n^{\frac{4-p_0}{2p_0}\times\frac{p_0-1}{2}}|\tilde{\sigma}^n(r,x_{\kappa(n,r)}^n)|^{p_0-1}\,dr\right)^{\frac{p_0}{p_0-1}} ds\\
&+ C + C\int_0^t \sup_{0\le r\le s}\mathbb{E}|x_r^n|^{p_0}ds,
\end{aligned}$$

for any $t\in[0,T]$. By using Hölder's inequality and Corollary 2.10,

$$\begin{aligned}
G_{52}(t) \le\, & Cn^{\frac{3p_0-4}{4(p_0-2)}-\frac{1}{p_0-1}}\int_0^t \mathbb{E}\int_{\kappa(n,s)}^s (1+|x_r^n|^{p_0}+|x_{\kappa(n,s)}^n|^{p_0})\,dr\,ds\\
&+ Cn^{-\frac{p_0}{2}+1-\frac{1}{p_0-1}}\int_0^t \mathbb{E}\int_{\kappa(n,s)}^s |\tilde{b}^n(r,x_{\kappa(n,r)}^n)|^{p_0}\,dr\,ds\\
&+ Cn^{\frac{3p_0-8}{4(p_0-3)}-\frac{1}{p_0-1}}\int_0^t \mathbb{E}\int_{\kappa(n,s)}^s (1+|x_r^n|^{p_0}+|x_{\kappa(n,s)}^n|^{p_0})\,dr\,ds\\
&+ Cn^{-\frac{p_0}{4}+1-\frac{1}{p_0-1}}\int_0^t \mathbb{E}\int_{\kappa(n,s)}^s |\tilde{\sigma}^n(r,x_{\kappa(n,r)}^n)|^{p_0}\,dr\,ds\\
&+ C + C\int_0^t \sup_{0\le r\le s}\mathbb{E}|x_r^n|^{p_0}ds,
\end{aligned}$$

for any $t\in[0,T]$. One observes that $n^{\frac{3p_0-4}{4(p_0-2)}}$ and $n^{\frac{3p_0-8}{4(p_0-3)}}$ are less than n for all $p_0\ge 4$, then due to Corollary 2.10, the following holds

$$G_{52}(t) \le C + C\int_0^t \sup_{0\le r\le s}\mathbb{E}|x_r^n|^{p_0}ds, \tag{2.6}$$

for any $t\in[0,T]$. In addition, note that by the definition of $b_2^n(t,x)$, one obtains

$$G_{53}(t) := p_0\mathbb{E}\int_0^t |x_{\kappa(n,s)}^n|^{p_0-2}\left\langle x_{\kappa(n,s)}^n, b_2^n(s,x_{\kappa(n,s)}^n)\right\rangle ds = 0, \tag{2.7}$$

for any $t \in [0,T]$. Then, substituting (2.5), (2.6) and (2.7) into (2.8), one obtains

$$G_5(t) \leq C + C\int_0^t \sup_{0\leq r\leq s} \mathbb{E}|x_r^n|^{p_0} ds, \tag{2.8}$$

for any $t \in [0,T]$. In order to estimate $G_6(t)$, one applies Young's inequality to obtain

$$\begin{aligned}
G_6(t) \leq & C\mathbb{E}\int_0^t |x_s^n|^{p_0}\, ds + C\mathbb{E}\int_0^t |\sigma_M^n(s, x_{\kappa(n,s)}^n)|^{p_0}\, ds \\
\leq & C\mathbb{E}\int_0^t |x_s^n|^{p_0}\, ds + C\mathbb{E}\int_0^t |\sigma_1^n(s, x_{\kappa(n,s)}^n)|^{p_0}\, ds \\
& + C\mathbb{E}\int_0^t |\sigma_2^n(s, x_{\kappa(n,s)}^n)|^{p_0}\, ds + C\mathbb{E}\int_0^t |\sigma_3^n(s, x_{\kappa(n,s)}^n)|^{p_0}\, ds,
\end{aligned}$$

which implies due to Lemma 2.9

$$G_6(t) \leq C + C\int_0^t \sup_{0\leq r\leq s} \mathbb{E}|x_r^n|^{p_0} ds,$$

for any $t \in [0,T]$. Finally, for $G_7(t)$, one writes

$$G_7(t) = \sum_{i=1}^{2} G_{7i}(t), \tag{2.9}$$

where

$$G_{71}(t) = p_0(p_0-1)\mathbb{E}\int_0^t (|x_s^n|^{p_0-2} - |x_{\kappa(n,s)}^n|^{p_0-2}) \sum_{k=1}^{d}\sum_{v=1}^{m} \sigma^{n,(k,v)}(x_{\kappa(n,s)}^n)\sigma_M^{n,(k,v)}(s, x_{\kappa(n,s)}^n)\, ds,$$

$$G_{72}(t) = p_0(p_0-1)\mathbb{E}\int_0^t |x_{\kappa(n,s)}^n|^{p_0-2} \sum_{k=1}^{d}\sum_{v=1}^{m} \sigma^{n,(k,v)}(x_{\kappa(n,s)}^n)\sigma_M^{n,(k,v)}(s, x_{\kappa(n,s)}^n)\, ds.$$

To estimate $G_{71}(t)$, one uses Itô's formula to obtain,

$$\begin{aligned}
& G_{71}(t) \\
:= & \ C\mathbb{E}\int_0^t (|x_s^n|^{p_0-2} - |x_{\kappa(n,s)}^n|^{p_0-2}) \sum_{k=1}^{d}\sum_{v=1}^{m} \sigma^{n,(k,v)}(x_{\kappa(n,s)}^n)\sigma_M^{n,(k,v)}(s, x_{\kappa(n,s)}^n)\, ds \\
\leq & \ C\mathbb{E}\int_0^t \int_{\kappa(n,s)}^s |x_r^n|^{p_0-4} \left\langle x_r^n, \tilde{b}^n(r, x_{\kappa(n,r)}^n)\right\rangle dr \sum_{k=1}^{d}\sum_{v=1}^{m} \sigma^{n,(k,v)}(x_{\kappa(n,s)}^n)\sigma_M^{n,(k,v)}(s, x_{\kappa(n,s)}^n)\, ds \\
& + C\mathbb{E}\int_0^t \int_{\kappa(n,s)}^s |x_r^n|^{p_0-4} \left\langle x_r^n, \tilde{\sigma}^n(r, x_{\kappa(n,r)}^n)\, dw_r\right\rangle \sum_{k=1}^{d}\sum_{v=1}^{m} \sigma^{n,(k,v)}(x_{\kappa(n,s)}^n)\sigma_M^{n,(k,v)}(s, x_{\kappa(n,s)}^n)\, ds \\
& + C\mathbb{E}\int_0^t \int_{\kappa(n,s)}^s |x_r^n|^{p_0-4} |\tilde{\sigma}^n(r, x_{\kappa(n,r)}^n)|^2\, dr \left|\sum_{k=1}^{d}\sum_{v=1}^{m} \sigma^{n,(k,v)}(x_{\kappa(n,s)}^n)\sigma_M^{n,(k,v)}(s, x_{\kappa(n,s)}^n)\right| ds,
\end{aligned}$$

which by using Remark 2.4 implies

$$\begin{aligned}
& G_{71}(t) \\
\leq & \ Cn^{\frac{1}{4}}\mathbb{E}\int_0^t \int_{\kappa(n,s)}^s |x_r^n|^{p_0-3}(1+|x_{\kappa(n,s)}^n|)|\tilde{b}^n(r, x_{\kappa(n,r)}^n)|\, dr |\sigma_M^n(s, x_{\kappa(n,s)}^n)|\, ds \\
& + C\mathbb{E}\int_0^t \int_{\kappa(n,s)}^s |x_r^n|^{p_0-4} \left\langle x_r^n, \tilde{\sigma}^n(r, x_{\kappa(n,r)}^n)\, dw_r\right\rangle
\end{aligned}$$

$$
\begin{aligned}
&\qquad\times \sum_{k=1}^{d}\sum_{v=1}^{m}\sigma^{n,(k,v)}(x^n_{\kappa(n,s)})\sum_{j=1}^{m}\int_{\kappa(n,s)}^{s}L^{n,j}\sigma^{(k,v)}(x^n_{\kappa(n,r)})\,dw^j_r\,ds\\
&+C\mathbb{E}\int_0^t\int_{\kappa(n,s)}^{s}|x^n_r|^{p_0-4}\left\langle x^n_r,\tilde{\sigma}^n(r,x^n_{\kappa(n,r)})\,dw_r\right\rangle\\
&\qquad\times \sum_{k=1}^{d}\sum_{v=1}^{m}\sigma^{n,(k,v)}(x^n_{\kappa(n,s)})\int_{\kappa(n,s)}^{s}L^{n,0}\sigma^{(k,v)}(x^n_{\kappa(n,r)})\,dr\,ds\\
&+C\mathbb{E}\int_0^t\int_{\kappa(n,s)}^{s}|x^n_r|^{p_0-4}\left\langle x^n_r,\tilde{\sigma}^n(r,x^n_{\kappa(n,r)})\,dw_r\right\rangle\\
&\qquad\times \sum_{k=1}^{d}\sum_{v=1}^{m}\sigma^{n,(k,v)}(x^n_{\kappa(n,s)})\sum_{j=1}^{m}\sum_{j_1=1}^{m}\int_{\kappa(n,s)}^{s}\int_{\kappa(n,r)}^{r}L^{n,j}L^{j_1}\sigma^{(k,v)}(x^n_{\kappa(n,r_1)})\,dw^{j_1}_{r_1}\,dw^j_r\,ds\\
&+Cn^{\frac14}\mathbb{E}\int_0^t\int_{\kappa(n,s)}^{s}|x^n_r|^{p_0-4}(1+|x^n_{\kappa(n,s)}|)|\tilde{\sigma}^n(r,x^n_{\kappa(n,r)})|^2\,dr|\sigma^n_M(s,x^n_{\kappa(n,s)})|\,ds,
\end{aligned}
$$

for any $t\in[0,T]$. One then observes that, since $L^{n,0}\sigma(x^n_{\kappa(n,r)})$ takes the same value for all $r\in[\kappa(n,s),s]$, it can be taken out of the integral in the third term above, and thus the third term is zero. Moreover, by Young's inequality and Remark 2.4, one obtains

$$
\begin{aligned}
&G_{71}(t)\\
&\le C\mathbb{E}\int_0^t\int_{\kappa(n,s)}^{s}n^{\frac14}(1+|x^n_r|^{p_0-2}+|x^n_{\kappa(n,s)}|^{p_0-2})|\tilde{b}^n(r,x^n_{\kappa(n,r)})|\,dr|\sigma^n_M(s,x^n_{\kappa(n,s)})|\,ds\\
&\quad+C\mathbb{E}\int_0^t\int_{\kappa(n,s)}^{s}n^{\frac34}|x^n_r|^{p_0-3}(1+|x^n_{\kappa(n,s)}|)^2|\tilde{\sigma}^n(r,x^n_{\kappa(n,r)})|\,dr\,ds\\
&\quad+C\sum_{j=1}^{m}\mathbb{E}\int_0^t\int_{\kappa(n,s)}^{s}n^{\frac34-\frac{2}{p_0}}|x^n_r|^{p_0-3}(1+|x^n_{\kappa(n,s)}|)n^{-\frac14+\frac{1}{p_0}}|\tilde{\sigma}^n(r,x^n_{\kappa(n,r)})|\\
&\qquad\qquad\times n^{-\frac14+\frac{1}{p_0}}\left|\sum_{j_1=1}^{d}\int_{\kappa(n,r)}^{r}L^{n,j}L^{j_1}\sigma(x^n_{\kappa(n,r_1)})\,dw^{j_1}_{r_1}\right|\,dr\,ds\\
&\quad+C\mathbb{E}\int_0^t\int_{\kappa(n,s)}^{s}n^{\frac14}(1+|x^n_r|^{p_0-3}+|x^n_{\kappa(n,s)}|^{p_0-3})|\tilde{\sigma}^n(r,x^n_{\kappa(n,r)})|^2\,dr|\sigma^{n,(i,j)}_M(s,x^n_{\kappa(n,s)})|\,ds,
\end{aligned}
$$

which yields, due to Young's inequality,

$$
\begin{aligned}
&G_{71}(t)\\
&\le C\int_0^t\mathbb{E}\left(\int_{\kappa(n,s)}^{s}n^{\frac34-\frac{1}{p_0}}(1+|x^n_r|^{p_0-2}+|x^n_{\kappa(n,s)}|^{p_0-2})n^{-\frac12+\frac{1}{p_0}}|\tilde{b}^n(r,x^n_{\kappa(n,r)})|\,dr\right)^{\frac{p_0}{p_0-1}}ds\\
&\quad+C\mathbb{E}\int_0^t\int_{\kappa(n,s)}^{s}n^{1-\frac{1}{p_0}}(1+|x^n_r|^{p_0-1}+|x^n_{\kappa(n,s)}|^{p_0-1})n^{-\frac14+\frac{1}{p_0}}|\tilde{\sigma}^n(r,x^n_{\kappa(n,r)})|\,dr\,ds\\
&\quad+C\mathbb{E}\int_0^t\int_{\kappa(n,s)}^{s}\left(n^{\frac34-\frac{2}{p_0}}(1+|x^n_r|^{p_0-2}+|x^n_{\kappa(n,s)}|^{p_0-2})n^{-\frac14+\frac{1}{p_0}}|\tilde{\sigma}^n(r,x^n_{\kappa(n,r)})|\right)^{\frac{p_0}{p_0-1}}dr\,ds\\
&\quad+C\int_0^t\mathbb{E}\left(\int_{\kappa(n,s)}^{s}n^{\frac34-\frac{2}{p_0}}(1+|x^n_r|^{p_0-3}+|x^n_{\kappa(n,s)}|^{p_0-3})n^{-\frac12+\frac{2}{p_0}}|\tilde{\sigma}^n(r,x^n_{\kappa(n,r)})|^2\,dr\right)^{\frac{p_0}{p_0-1}}ds\\
&\quad+C\sum_{j=1}^{m}n^{-\frac{p_0}{4}+1}\mathbb{E}\int_0^t\int_{\kappa(n,s)}^{s}\left|\sum_{j_1=1}^{d}\int_{\kappa(n,r)}^{r}L^{n,j}L^{j_1}\sigma(x^n_{\kappa(n,r_1)})\,dw^{j_1}_{r_1}\right|^{p_0}dr\,ds\\
&\quad+C\int_0^t\mathbb{E}|\sigma^n_M(s,x^n_{\kappa(n,s)})|^{p_0}\,ds,
\end{aligned}
$$

for any $t \in [0,T]$. By Young's inequality, Hölder's inequality and Lemma 2.9,

$$
\begin{aligned}
G_{71}(t) \leq & Cn^{\frac{3p_0-4}{4(p_0-2)}-\frac{1}{p_0-1}} \int_0^t \mathbb{E}\int_{\kappa(n,s)}^s (1+|x_r^n|^{p_0}+|x_{\kappa(n,s)}^n|^{p_0})\,dr\,ds \\
&+ Cn^{-\frac{p_0}{2}+1-\frac{1}{p_0-1}} \int_0^t \int_{\kappa(n,s)}^s \mathbb{E}|\tilde{b}^n(r,x_{\kappa(n,r)}^n)|^{p_0}\,dr\,ds \\
&+ Cn\mathbb{E}\int_0^t \int_{\kappa(n,s)}^s (1+|x_r^n|^{p_0}+|x_{\kappa(n,s)}^n|^{p_0})\,dr\,ds \\
&+ Cn^{-\frac{p_0}{4}+1} \int_0^t \int_{\kappa(n,s)}^s \mathbb{E}|\tilde{\sigma}^n(r,x_{\kappa(n,r)}^n)|^{p_0}\,dr\,ds \\
&+ Cn^{\frac{3p_0-8}{4(p_0-2)}}\mathbb{E}\int_0^t \int_{\kappa(n,s)}^s (1+|x_r^n|^{p_0}+|x_{\kappa(n,s)}^n|^{p_0})\,dr\,ds \\
&+ Cn^{-\frac{p_0}{4}+1} \int_0^t \int_{\kappa(n,s)}^s \mathbb{E}|\tilde{\sigma}^n(r,x_{\kappa(n,r)}^n)|^{p_0}\,dr\,ds \\
&+ Cn^{\frac{3p_0-8}{4(p_0-3)}-\frac{1}{p_0-1}} \int_0^t \mathbb{E}\int_{\kappa(n,s)}^s (1+|x_r^n|^{p_0}+|x_{\kappa(n,s)}^n|^{p_0})\,dr\,ds \\
&+ Cn^{-\frac{p_0}{4}+1-\frac{1}{p_0-1}} \int_0^t \int_{\kappa(n,s)}^s \mathbb{E}|\tilde{\sigma}^n(r,x_{\kappa(n,r)}^n)|^{p_0}\,dr\,ds \\
&+ Cn^{-\frac{3p_0}{4}+2} \int_0^t \int_{\kappa(n,s)}^s \int_{\kappa(n,r)}^r \mathbb{E}|L^{n,j}L^{j_1}\sigma(x_{\kappa(n,r_1)}^n)|^{p_0}\,dr_1\,dr\,ds \\
&+ C + C\int_0^t \sup_{0\leq r\leq s} \mathbb{E}|x_r^n|^{p_0} ds,
\end{aligned}
$$

for any $t \in [0,T]$. Due to Corollary 2.10 and Remark 2.4, it can be shown that

$$
G_{71}(t) \leq C + C\int_0^t \sup_{0\leq r\leq s} \mathbb{E}|x_r^n|^{p_0} ds, \tag{2.10}
$$

for any $t \in [0,T]$. In order to estimate $G_{72}(t)$, one writes

$$
\begin{aligned}
G_{72}(t) = & \, p_0(p_0-1)\mathbb{E}\int_0^t |x_{\kappa(n,s)}^n|^{p_0-2} \sum_{k=1}^d \sum_{v=1}^m \sigma^{n,(k,v)}(x_{\kappa(n,s)}^n) \\
& \qquad \times \sum_{j=1}^m \int_{\kappa(n,s)}^s L^{n,j}\sigma^{(k,v)}(x_{\kappa(n,r)}^n)\,dw_r^j\,ds \\
&+ p_0(p_0-1)\mathbb{E}\int_0^t |x_{\kappa(n,s)}^n|^{p_0-2} \sum_{k=1}^d \sum_{v=1}^m \sigma^{n,(k,v)}(x_{\kappa(n,s)}^n) \\
& \qquad \times \int_{\kappa(n,s)}^s L^{n,0}\sigma^{(k,v)}(x_{\kappa(n,r)}^n)\,dr\,ds \\
&+ p_0(p_0-1)\mathbb{E}\int_0^t |x_{\kappa(n,s)}^n|^{p_0-2} \sum_{k=1}^d \sum_{v=1}^m \sigma^{n,(k,v)}(x_{\kappa(n,s)}^n) \\
& \qquad \times \sum_{j=1}^m \sum_{j_1=1}^m \int_{\kappa(n,s)}^s \int_{\kappa(n,r)}^r L^{n,j}L^{j_1}\sigma^{(k,v)}(x_{\kappa(n,r_1)}^n)\,dw_{r_1}^{j_1}\,dw_r^j\,ds,
\end{aligned}
$$

which implies, due to Remark 2.4 and the fact that the first and third terms are zero,

$$
G_{72}(t) \leq C\mathbb{E}\int_0^t \int_{\kappa(n,s)}^s n(1+|x_{\kappa(n,s)}^n|)^{p_0}\,dr\,ds,
$$

for any $t \in [0,T]$. Then, one obtains

$$G_{72}(t) \leq C + C\int_0^t \sup_{0\leq r\leq s} \mathbb{E}|x_r^n|^{p_0} ds, \tag{2.11}$$

for any $t \in [0,T]$. Furthermore, substituting (2.10) and (2.11) into (2.9) yields

$$G_7 \leq C + C\int_0^t \sup_{0\leq r\leq s} \mathbb{E}|x_r^n|^{p_0} ds,$$

for any $t \in [0,T]$. Therefore, for any $n \in \mathbb{N}$ and $t \in [0,T]$,

$$\sup_{0\leq s\leq t} \mathbb{E}|x_s^n|^{p_0} \leq C + C\int_0^t \sup_{0\leq r\leq s} \mathbb{E}|x_r^n|^{p_0} ds < \infty,$$

and applying Gronwall's lemma completes the proof. □

2.4 Proof of main result

Lemma 2.12. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a twice continuously differentiable function. If there exist constants $\bar{\alpha} \in \mathbb{R}$, $K > 0$ and $\alpha \in (0,1]$, such that for any $x, \bar{x} \in \mathbb{R}^d$,*

$$|\nabla^2 f(x) - \nabla^2 f(\bar{x})| \leq K(1 + |x| + |\bar{x}|)^{\bar{\alpha}}|x - \bar{x}|^{\alpha},$$

then, there is a constant $C > 0$ such that for any $x, \bar{x} \in \mathbb{R}^d$, and $i = 1, \ldots, d$,

$$\left|\frac{\partial f(x)}{\partial y^{(i)}} - \frac{\partial f(\bar{x})}{\partial y^{(i)}} - \sum_{j=1}^d \frac{\partial^2 f(\bar{x})}{\partial y^{(i)}\partial y^{(j)}}(x^{(j)} - \bar{x}^{(j)})\right| \leq C(1 + |x| + |\bar{x}|)^{\bar{\alpha}}|x - \bar{x}|^{1+\alpha}.$$

Proof. One uses the mean value theorem to obtain that, for all $x, \bar{x} \in \mathbb{R}^d$, $i = 1, \ldots, d$, there exists $q \in [0,1]$, such that

$$\frac{\partial f(x)}{\partial y^{(i)}} - \frac{\partial f(\bar{x})}{\partial y^{(i)}} = \sum_{j=1}^d \frac{\partial^2 f((qx + (1-q)\bar{x})}{\partial y^{(i)}\partial y^{(j)}}(x^{(j)} - \bar{x}^{(j)}).$$

Then for a fixed $q \in (0,1)$,

$$\begin{aligned}
&\left|\frac{\partial f(x)}{\partial y^{(i)}} - \frac{\partial f(\bar{x})}{\partial y^{(i)}} - \sum_{j=1}^d \frac{\partial^2 f(\bar{x})}{\partial y^{(i)}\partial y^{(j)}}(x^{(j)} - \bar{x}^{(j)})\right| \\
&= \left|\sum_{j=1}^d \frac{\partial^2 f((qx + (1-q)\bar{x})}{\partial y^{(i)}\partial y^{(j)}}(x^{(j)} - \bar{x}^{(j)}) - \sum_{j=1}^d \frac{\partial^2 f(\bar{x})}{\partial y^{(i)}\partial y^{(j)}}(x^{(j)} - \bar{x}^{(j)})\right| \\
&\leq \sum_{j=1}^d \left|\frac{\partial^2 f((qx + (1-q)\bar{x})}{\partial y^{(i)}\partial y^{(j)}} - \frac{\partial^2 f(\bar{x})}{\partial y^{(i)}\partial y^{(j)}}\right| |x^{(j)} - \bar{x}^{(j)}| \\
&\leq C(1 + |x| + |\bar{x}|)^{\bar{\alpha}}|x - \bar{x}|^{1+\alpha}.
\end{aligned}$$

□

Lemma 2.13. *Assume **A-1** to **A-5** hold, then, there exists a constant $C > 0$, such that for any $p \leq \frac{p_0}{2\rho+1}$ and $n \in \mathbb{N}$,*

$$\sup_{0\leq t\leq T} \mathbb{E}|b_1^n(t, x_{\kappa(n,t)}^n)|^p \leq Cn^{-p}, \quad \sup_{0\leq t\leq T} \mathbb{E}|b_2^n(t, x_{\kappa(n,t)}^n)|^p \leq Cn^{-\frac{p}{2}},$$

$$\sup_{0\le t\le T}\mathbb{E}|\sigma_1^n(t,x^n_{\kappa(n,t)})|^p\le Cn^{-\frac{p}{2}},\quad \sup_{0\le t\le T}\mathbb{E}|\sigma_2^n(t,x^n_{\kappa(n,t)})|^p\le Cn^{-p},$$

$$\sup_{0\le t\le T}\mathbb{E}|\sigma_3^n(t,x^n_{\kappa(n,t)})|^p\le Cn^{-p}.$$

Proof. By applying Hölder's inequality and Remark 2.3, one obtains, for any $p\le\frac{p_0}{2\rho+1}$,

$$\begin{aligned}\mathbb{E}|b_1^n(t,x^n_{\kappa(n,t)})|^p&=\mathbb{E}\left|\int_{\kappa(n,t)}^t L^{n,0}b(x^n_{\kappa(n,s)})\,ds\right|^p\\&\le Cn^{-p+1}\int_{\kappa(n,t)}^t\mathbb{E}|L^{n,0}b(x^n_{\kappa(n,s)})|^p\,ds\\&\le Cn^{-p+1}\int_{\kappa(n,t)}^t\mathbb{E}(1+|x^n_{\kappa(n,s)}|)^{(2\rho+1)p}\,ds\\&\le Cn^{-p},\end{aligned}$$

where the last inequality holds due to Lemma 2.11. Other results can be proved by using similar arguments. ☐

Corollary 2.14. *Assume **A-1** to **A-5** hold, then, there exists a constant $C>0$, such that for any $p\le\frac{p_0}{2\rho+1}$ and $n\in\mathbb{N}$,*

$$\sup_{0\le t\le T}\mathbb{E}|\tilde{b}^n(t,x^n_{\kappa(n,t)})|^p\le C,\quad \sup_{0\le t\le T}\mathbb{E}|\tilde{\sigma}^n(t,x^n_{\kappa(n,t)})|^p\le C.$$

Lemma 2.15. *Assume **A-1** to **A-5** hold, then, there exists a constant $C>0$, such that for any $p\le\frac{p_0}{2\rho+1}$ and $n\in\mathbb{N}$,*

$$\sup_{0\le t\le T}\mathbb{E}|x^n_t-x^n_{\kappa(n,t)}|^p\le Cn^{-\frac{p}{2}}.$$

Proof. For $p\ge1$, by using Hölder's inequality, one obtains

$$\begin{aligned}\mathbb{E}|x^n_t-x^n_{\kappa(n,t)}|^p&\le C\mathbb{E}\left|\int_{\kappa(n,t)}^t\tilde{b}^n(s,x^n_{\kappa(n,s)})\,ds\right|^p+C\mathbb{E}\left|\int_{\kappa(n,t)}^t\tilde{\sigma}^n(s,x^n_{\kappa(n,s)})\,dw_s\right|^p\\&\le n^{-p+1}C\mathbb{E}\int_{\kappa(n,t)}^t|\tilde{b}^n(s,x^n_{\kappa(n,s)})|^p\,ds+Cn^{-\frac{p}{2}+1}\mathbb{E}\int_{\kappa(n,t)}^t|\tilde{\sigma}^n(s,x^n_{\kappa(n,s)})|^p\,ds,\end{aligned}$$

which by using corollary 2.14 yields the desired result. As for $p\in(0,1)$, one uses Jensen's inequality to obtain the same result. ☐

Lemma 2.16. *Assume **A-1** to **A-5** hold, then, there exists a constant $C>0$, such that for any $p\le\frac{p_0}{2\rho+1}$ and $n\in\mathbb{N}$,*

$$\sup_{0\le t\le T}\mathbb{E}|b(x^n_{\kappa(n,t)})-b^n(x^n_{\kappa(n,t)})|^p\le Cn^{-\frac{3}{2}p},\quad \sup_{0\le t\le T}\mathbb{E}|\sigma(x^n_{\kappa(n,t)})-\sigma^n(x^n_{\kappa(n,t)})|^p\le Cn^{-\frac{3}{2}p}$$

Proof. We have the following expression,

$$|b(x^n_{\kappa(n,t)})-b^n(x^n_{\kappa(n,t)})|=n^{-\frac{3}{2}}\frac{|x^n_{\kappa(n,t)}|^{3\rho}|b(x^n_{\kappa(n,t)})|}{1+n^{-\frac{3}{2}}|x^n_{\kappa(n,t)}|^{3\rho}}\le n^{-\frac{3}{2}}(1+|x^n_{\kappa(n,t)}|)^{4\rho+1},$$

and then by using Lemma 2.11 and the same argument for σ completes the proof. ☐

Lemma 2.17. *Assume **A-1** to **A-5** hold and $p_0\ge2(5\rho+1)$. Then, there exists a constant $C>0$, such that for any $n\in\mathbb{N}$,*

$$\sup_{0\le t\le T}\mathbb{E}|\sigma(x^n_t)-\sigma(x^n_{\kappa(n,t)})-\sigma^n_M(t,x^n_{\kappa(n,t)})|^2\le Cn^{-(2+\alpha)}.$$

Proof. For every $k=1,\ldots,d$, $v=1,\ldots,m$, applying Itô's formula to $\sigma^{(k,v)}(x_t^n)-\sigma^{(k,v)}(x^n_{\kappa(n,t)})$ gives, almost surely,

$$
\begin{aligned}
&\sigma^{(k,v)}(x_t^n)-\sigma^{(k,v)}(x^n_{\kappa(n,t)}) \\
&=\sum_{i=1}^{d}\int_{\kappa(n,t)}^{t}\frac{\partial\sigma^{(k,v)}(x_s^n)}{\partial x^{(i)}}\tilde{b}^{n,(i)}(s,x^n_{\kappa(n,s)})\,ds+\sum_{i=1}^{d}\sum_{j=1}^{m}\int_{\kappa(n,t)}^{t}\frac{\partial\sigma^{(k,v)}(x_s^n)}{\partial x^{(i)}}\tilde{\sigma}^{n,(i,j)}(s,x^n_{\kappa(n,s)})\,dw_s^j \\
&\quad+\frac{1}{2}\sum_{i,l=1}^{d}\sum_{j=1}^{m}\int_{\kappa(n,t)}^{t}\frac{\partial^2\sigma^{(k,v)}(x_s^n)}{\partial x^{(i)}\partial x^{(l)}}\tilde{\sigma}^{n,(i,j)}(s,x^n_{\kappa(n,s)})\tilde{\sigma}^{n,(l,j)}(s,x^n_{\kappa(n,s)})\,ds=\sum_{i=1}^{12}J_i(t)
\end{aligned}
$$

where

$$
\begin{aligned}
J_1(t)&=\sum_{i=1}^{d}\int_{\kappa(n,t)}^{t}\left(\frac{\partial\sigma^{(k,v)}(x_s^n)}{\partial x^{(i)}}-\frac{\partial\sigma^{(k,v)}(x^n_{\kappa(n,s)})}{\partial x^{(i)}}\right)b^{n,(i)}(x^n_{\kappa(n,s)})\,ds,\\
J_2(t)&=\sum_{i=1}^{d}\int_{\kappa(n,t)}^{t}\frac{\partial\sigma^{(k,v)}(x^n_{\kappa(n,s)})}{\partial x^{(i)}}b^{n,(i)}(x^n_{\kappa(n,s)})\,ds,\\
J_3(t)&=\sum_{i=1}^{d}\int_{\kappa(n,t)}^{t}\frac{\partial\sigma^{(k,v)}(x_s^n)}{\partial x^{(i)}}(b_1^{n,(i)}(s,x^n_{\kappa(n,s)})+b_2^{n,(i)}(s,x^n_{\kappa(n,s)}))\,ds,\\
J_4(t)&=\sum_{i=1}^{d}\sum_{j=1}^{m}\int_{\kappa(n,t)}^{t}\Bigg(\frac{\partial\sigma^{(k,v)}(x_s^n)}{\partial x^{(i)}}-\frac{\partial\sigma^{(k,v)}(x^n_{\kappa(n,s)})}{\partial x^{(i)}}\\
&\qquad-\sum_{l=1}^{d}\frac{\partial^2\sigma^{(k,v)}(x^n_{\kappa(n,s)})}{\partial x^{(i)}\partial x^{(l)}}(x_s^{n,(l)}-x^{n,(l)}_{\kappa(n,s)})\Bigg)\sigma^{n,(i,j)}(x^n_{\kappa(n,s)})\,dw_s^j,\\
J_5(t)&=\sum_{i=1}^{d}\sum_{j=1}^{m}\int_{\kappa(n,t)}^{t}\sum_{l=1}^{d}\frac{\partial^2\sigma^{(k,v)}(x^n_{\kappa(n,s)})}{\partial x^{(i)}\partial x^{(l)}}\Bigg(\int_{\kappa(n,s)}^{s}\tilde{b}^{n,(l)}(r,x^n_{\kappa(n,r)})\,dr\\
&\qquad+\sum_{j_1=1}^{m}\int_{\kappa(n,s)}^{s}\sigma_M^{n,(l,j_1)}(r,x^n_{\kappa(n,r)})\,dw_r^{j_1}\Bigg)\sigma^{n,(i,j)}(x^n_{\kappa(n,s)})\,dw_s^j,\\
J_6(t)&=\sum_{i=1}^{d}\sum_{j=1}^{m}\int_{\kappa(n,t)}^{t}\left(\sum_{l=1}^{d}\frac{\partial^2\sigma^{(k,v)}(x^n_{\kappa(n,s)})}{\partial x^{(i)}\partial x^{(l)}}\sum_{j_1=1}^{m}\int_{\kappa(n,s)}^{s}\sigma^{n,(l,j_1)}(x^n_{\kappa(n,r)})\,dw_r^{j_1}\right)\\
&\qquad\times\sigma^{n,(i,j)}(x^n_{\kappa(n,s)})\,dw_s^j,\\
J_7(t)&=\sum_{i=1}^{d}\sum_{j=1}^{m}\int_{\kappa(n,t)}^{t}\left(\frac{\partial\sigma^{(k,v)}(x_s^n)}{\partial x^{(i)}}-\frac{\partial\sigma^{(k,v)}(x^n_{\kappa(n,s)})}{\partial x^{(i)}}\right)\sigma_1^{n,(i,j)}(s,x^n_{\kappa(n,s)})\,dw_s^j,\\
J_8(t)&=\sum_{i=1}^{d}\sum_{j=1}^{m}\int_{\kappa(n,t)}^{t}\frac{\partial\sigma^{(k,v)}(x^n_{\kappa(n,s)})}{\partial x^{(i)}}(\sigma^{n,(i,j)}(x^n_{\kappa(n,s)})+\sigma_1^{n,(i,j)}(s,x^n_{\kappa(n,s)}))\,dw_s^j,\\
J_9(t)&=\sum_{i=1}^{d}\sum_{j=1}^{m}\int_{\kappa(n,t)}^{t}\frac{\partial\sigma^{(k,v)}(x_s^n)}{\partial x^{(i)}}(\sigma_2^{n,(i,j)}(s,x^n_{\kappa(n,s)})+\sigma_3^{n,(i,j)}(s,x^n_{\kappa(n,s)}))\,dw_s^j,\\
J_{10}(t)&=\frac{1}{2}\sum_{i,l=1}^{d}\sum_{j=1}^{m}\int_{\kappa(n,t)}^{t}\left(\frac{\partial^2\sigma^{(k,v)}(x_s^n)}{\partial x^{(i)}\partial x^{(l)}}-\frac{\partial^2\sigma^{(k,v)}(x^n_{\kappa(n,s)})}{\partial x^{(i)}\partial x^{(l)}}\right)\\
&\qquad\times\sigma^{n,(i,j)}(x^n_{\kappa(n,s)})\sigma^{n,(l,j)}(x^n_{\kappa(n,s)})\,ds,\\
J_{11}(t)&=\frac{1}{2}\sum_{i,l=1}^{d}\sum_{j=1}^{m}\int_{\kappa(n,t)}^{t}\frac{\partial^2\sigma^{(k,v)}(x^n_{\kappa(n,s)})}{\partial x^{(i)}\partial x^{(l)}}\sigma^{n,(i,j)}(x^n_{\kappa(n,s)})\sigma^{n,(l,j)}(x^n_{\kappa(n,s)})\,ds,
\end{aligned}
$$

$$J_{12}(t)=\frac{1}{2}\sum_{i,l=1}^{d}\sum_{j=1}^{m}\int_{\kappa(n,t)}^{t}\frac{\partial^2\sigma^{(k,v)}(x_s^n)}{\partial x^{(i)}\partial x^{(l)}}(\sigma^{n,(i,j)}(x_{\kappa(n,s)}^n)\sigma_M^{n,(l,j)}(s,x_{\kappa(n,s)}^n)$$
$$+\sigma_M^{n,(i,j)}(s,x_{\kappa(n,s)}^n)\tilde{\sigma}^{n,(l,j)}(s,x_{\kappa(n,s)}^n))\,ds.$$

It can be observed that

$$\begin{aligned}
&\mathbb{E}|J_2(t)+J_6(t)+J_8(t)+J_{11}(t)-\sigma_M^{n,(k,v)}(t,x_{\kappa(n,t)}^n)|^2\\
&\le 2\mathbb{E}|J_2(t)+J_{11}(t)-\sigma_2^{n,(k,v)}(t,x_{\kappa(n,t)}^n)|^2\\
&\quad+2\mathbb{E}|J_6(t)+J_8(t)-\sigma_1^{n,(k,v)}(t,x_{\kappa(n,t)}^n)-\sigma_3^{n,(k,v)}(t,x_{\kappa(n,t)}^n)|^2\\
&\le C\sum_{i,l=1}^{d}\sum_{j=1}^{m}\mathbb{E}\Bigg|-\frac{n^{-3/2}|x_{\kappa(n,t)}^n|^{3\rho}}{(1+n^{-3/2}|x_{\kappa(n,t)}^n|^{3\rho})^2}\\
&\qquad\times\int_{\kappa(n,t)}^{t}\frac{\partial^2\sigma^{(k,v)}(x_{\kappa(n,s)}^n)}{\partial x^{(i)}\partial x^{(l)}}\sigma^{(i,j)}(x_{\kappa(n,s)}^n)\sigma^{(l,j)}(x_{\kappa(n,s)}^n)\,ds\Bigg|^2\\
&\quad+2\mathbb{E}\Bigg|-\frac{n^{-3/2}|x_{\kappa(n,t)}^n|^{3\rho}}{(1+n^{-3/2}|x_{\kappa(n,t)}^n|^{3\rho})^2}\sum_{i,l=1}^{d}\sum_{j,j_1=1}^{m}\int_{\kappa(n,t)}^{t}\frac{\partial^2\sigma^{(k,v)}(x_{\kappa(n,s)}^n)}{\partial x^{(i)}\partial x^{(l)}}\\
&\qquad\times\int_{\kappa(n,s)}^{s}\sigma^{(l,j_1)}(x_{\kappa(n,r)}^n)\,dw_r^{j_1}\sigma^{(i,j)}(x_{\kappa(n,s)}^n)\,dw_s^j\Bigg|^2,
\end{aligned}$$

which implies due to Remark 2.3 and Lemma 2.11 that

$$\begin{aligned}
&\mathbb{E}|J_2(t)+J_6(t)+J_8(t)+J_{11}(t)-\sigma_M^{n,(k,v)}(t,x_{\kappa(n,t)}^n)|^2\\
&\le Cn^{-3}\mathbb{E}|n^{-1}|x_{\kappa(n,t)}^n|^{3\rho}(1+|x_{\kappa(n,t)}^n|^{3/2\rho+1})|^2\\
&\quad+Cn^{-5}\mathbb{E}||x_{\kappa(n,t)}^n|^{3\rho}(1+|x_{\kappa(n,t)}^n|^{3/2\rho+1})|^2\le Cn^{-5},
\end{aligned}$$

for $p_0\ge 9\rho+2$. Then, one obtains the following

$$\begin{aligned}
&\mathbb{E}|\sigma^{(k,v)}(x_t^n)-\sigma^{(k,v)}(x_{\kappa(n,t)}^n)-\sigma_M^{n,(k,v)}(t,x_{\kappa(n,t)}^n)|^2\\
&\le 2\mathbb{E}|J_1(t)+J_3(t)+J_4(t)+J_5(t)+J_7(t)+J_9(t)+J_{10}(t)+J_{12}(t)|^2\\
&\quad+2\mathbb{E}|J_2(t)+J_6(t)+J_8(t)+J_{11}(t)-\sigma_M^{n,(k,v)}(t,x_{\kappa(n,t)}^n)|^2\\
&\le C(\mathbb{E}|J_1(t)|^2+\mathbb{E}|J_3(t)|^2+\mathbb{E}|J_4(t)|^2+\mathbb{E}|J_5(t)|^2+\mathbb{E}|J_7(t)|^2\\
&\quad+\mathbb{E}|J_9(t)|^2+\mathbb{E}|J_{10}(t)|^2+\mathbb{E}|J_{12}(t)|^2)+Cn^{-5},
\end{aligned}$$

for any $t\in[0,T]$. By using Cauchy-Schwarz inequality, $\mathbb{E}|J_1(t)|^2$ can be estimated as

$$\mathbb{E}|J_1(t)|^2\le Cn^{-1}\sum_{i=1}^{d}\int_{\kappa(n,t)}^{t}\mathbb{E}\left|\left(\frac{\partial\sigma^{(k,v)}(x_s^n)}{\partial x^{(i)}}-\frac{\partial\sigma^{(k,v)}(x_{\kappa(n,s)}^n)}{\partial x^{(i)}}\right)b^{n,(i)}(x_{\kappa(n,s)}^n)\right|^2 ds,$$

which by using Young's inequality, Remark 2.3 and Hölder's inequality yields

$$\begin{aligned}
\mathbb{E}|J_1(t)|^2&\le Cn^{-1}\int_{\kappa(n,t)}^{t}\mathbb{E}(1+|x_s^n|+|x_{\kappa(n,s)}^n|)^{\rho-2}(1+|x_{\kappa(n,s)}^n|)^{2\rho+2}|x_s^n-x_{\kappa(n,s)}^n|^2\,ds\\
&\le Cn^{-1}\int_{\kappa(n,t)}^{t}\mathbb{E}(1+|x_s^n|^{3\rho}+|x_{\kappa(n,s)}^n|^{3\rho})|x_s^n-x_{\kappa(n,s)}^n|^2\,ds\\
&\le Cn^{-1}\int_{\kappa(n,t)}^{t}\left(\mathbb{E}(1+|x_s^n|^{p_0}+|x_{\kappa(n,s)}^n|^{p_0})\right)^{\frac{3\rho}{p_0}}\left(\mathbb{E}|x_s^n-x_{\kappa(n,s)}^n|^{\frac{2p_0}{p_0-3\rho}}\right)^{\frac{p_0-3\rho}{p_0}}ds,
\end{aligned}$$

for any $t \in [0,T]$. One uses Lemma 2.11 and Lemma 2.15 to obtain

$$\mathbb{E}|J_1(t)|^2 \leq Cn^{-3},$$

for every $n \in \mathbb{N}$. To estimate $\mathbb{E}|J_3(t)|^2$, one applies Cauchy-Schwarz inequality and Remark 2.3 to obtain

$$\mathbb{E}|J_3(t)|^2 \leq Cn^{-1} \int_{\kappa(n,t)}^{t} \mathbb{E}(1+|x_s^n|)^{\rho}(|b_1^n(s,x_{\kappa(n,s)}^n)|^2 + |b_2^n(s,x_{\kappa(n,s)}^n)|^2)\, ds,$$

which implies due to Hölder's inequality

$$\mathbb{E}|J_3(t)|^2 \leq Cn^{-1} \int_{\kappa(n,t)}^{t} (\mathbb{E}(1+|x_s^n|^{p_0}))^{\frac{\rho}{p_0}} (\mathbb{E}(|b_1^n(s,x_{\kappa(n,s)}^n)|^{\frac{2p_0}{p_0-\rho}} + |b_2^n(s,x_{\kappa(n,s)}^n)|^{\frac{2p_0}{p_0-\rho}}))^{\frac{p_0-\rho}{p_0}}\, ds,$$

for any $t \in [0,T]$. By Lemma 2.11 and Lemma 2.13, it becomes

$$\mathbb{E}|J_3(t)|^2 \leq Cn^{-3},$$

for every $n \in \mathbb{N}$. As for $\mathbb{E}|J_4(t)|^2$, by using Young's inequality, Cauchy-Schwarz inequality, Remark 2.3 and Lemma 2.12, one obtains

$$\begin{aligned}\mathbb{E}|J_4(t)|^2 &\leq C\int_{\kappa(n,t)}^{t} \mathbb{E}((1+|x_s^n|+|x_{\kappa(n,s)}^n|)^{\rho-4}(1+|x_{\kappa(n,s)}^n|)^{\rho+2}|x_s^n - x_{\kappa(n,s)}^n|^{2+2\alpha}\, ds\\ &\leq C\int_{\kappa(n,t)}^{t} \mathbb{E}((1+|x_s^n|+|x_{\kappa(n,s)}^n|)^{2\rho-2}|x_s^n - x_{\kappa(n,s)}^n|^{2+2\alpha})\, ds,\end{aligned}$$

which implies due to Hölder's inequality

$$\mathbb{E}|J_4(t)|^2 \leq C\int_{\kappa(n,t)}^{t} \Big(\mathbb{E}(1+|x_s^n|^{p_0}+|x_{\kappa(n,s)}^n|^{p_0})\Big)^{\frac{2\rho-2}{p_0}} \Big(\mathbb{E}|x_s^n - x_{\kappa(n,s)}^n|^{\frac{(2+2\alpha)p_0}{p_0-2\rho+2}}\Big)^{\frac{p_0-2\rho+2}{p_0}}\, ds,$$

for any $t \in [0,T]$. Then, applying Lemma 2.15 and Lemma 2.11 yield

$$\mathbb{E}|J_4(t)|^2 \leq Cn^{-(2+\alpha)},$$

for every $n \in \mathbb{N}$. In order to estimate $\mathbb{E}|J_5(t)|^2$, one uses Young's inequality and Cauchy-Schwarz inequality to obtain

$$\begin{aligned}\mathbb{E}|J_5(t)|^2 \leq C\int_{\kappa(n,t)}^{t} \mathbb{E}\left|\int_{\kappa(n,s)}^{s} |\tilde{b}^n(r,x_{\kappa(n,r)}^n)|\, dr + \left|\sum_{j_1=1}^{m}\int_{\kappa(n,s)}^{s} \sigma_M^{n,(i,j_1)}(r,x_{\kappa(n,r)}^n)\, dw_r^{j_1}\right|\right|^2\\ \times (1+|x_{\kappa(n,s)}^n|)^{2\rho}\, ds,\end{aligned}$$

which, by applying Hölder's inequality, yields

$$\begin{aligned}\mathbb{E}|J_5(t)|^2 \leq C\int_{\kappa(n,t)}^{t} \Bigg(& n^{-\frac{2p_0}{p_0-2\rho}+1}\int_{\kappa(n,s)}^{s} \mathbb{E}|\tilde{b}^n(r,x_{\kappa(n,r)}^n)|^{\frac{2p_0}{p_0-2\rho}}\, ds\\ &+ n^{-\frac{p_0}{p_0-2\rho}+1}\int_{\kappa(n,s)}^{s} \mathbb{E}|\sigma_M^n(r,x_{\kappa(n,r)}^n)|^{\frac{2p_0}{p_0-2\rho}}\, ds\Bigg)^{\frac{p_0-2\rho}{p_0}} \Big(\mathbb{E}(1+|x_{\kappa(n,s)}^n|^{p_0})\Big)^{\frac{2\rho}{p_0}}\, ds,\end{aligned}$$

for any $t \in [0,T]$. One uses Corollary 2.14 and Lemma 2.13 to obtain

$$\mathbb{E}|J_5(t)|^2 \leq Cn^{-3},$$

for every $n \in \mathbb{N}$. As for $\mathbb{E}|J_7(t)|^2$, it can be estimated by using Cauchy-Schwarz inequality as follows

$$\mathbb{E}|J_7(t)|^2 \leq C\sum_{i=1}^{d}\sum_{j=1}^{m}\int_{\kappa(n,t)}^{t}\mathbb{E}\left|\frac{\partial\sigma^{(k,v)}(x_s^n)}{\partial x^{(i)}}-\frac{\partial\sigma^{(k,v)}(x_{\kappa(n,s)}^n)}{\partial x^{(i)}}\right|^2|\sigma_1^n(s,x_{\kappa(n,s)}^n)|^2\,ds,$$

which yields by using Remark 2.3 and Hölder's inequality

$$\begin{aligned}\mathbb{E}|J_7(t)|^2 &\leq C\int_{\kappa(n,t)}^{t}\mathbb{E}(1+|x_s^n|+|x_{\kappa(n,s)}^n|)^{\rho-2}|x_s^n-x_{\kappa(n,s)}^n|^2|\sigma_1^n(s,x_{\kappa(n,s)}^n)|^2\,ds\\ &\leq C\int_{\kappa(n,t)}^{t}\left(\mathbb{E}(1+|x_s^n|+|x_{\kappa(n,s)}^n|)^{p_0}\right)^{\frac{\rho-2}{p_0}}\\ &\qquad\times\left(\mathbb{E}|x_s^n-x_{\kappa(n,s)}^n)|^{\frac{2p_0}{p_0-\rho+2}}|\sigma_1^{n,}(s,x_{\kappa(n,s)}^n)|^{\frac{2p_0}{p_0-\rho+2}}\right)^{\frac{p_0-\rho+2}{p_0}}\,ds,\end{aligned}$$

for $\rho>2$, and any $t\in[0,T]$. Then, one can apply Cauchy-Schwarz inequality and Lemma 2.11 to obtain

$$\mathbb{E}|J_7(t)|^2 \leq C\int_{\kappa(n,t)}^{t}\left(\mathbb{E}|x_s^n-x_{\kappa(n,s)}^n)|^{\frac{4p_0}{p_0-\rho+2}}\mathbb{E}|\sigma_1^{n,}(s,x_{\kappa(n,s)}^n)|^{\frac{4p_0}{p_0-\rho+2}}\right)^{\frac{p_0-\rho+2}{2p_0}}\,ds,$$

Thus, applying Lemma 2.15 and Lemma 2.13 give the following estimate

$$\mathbb{E}|J_7(t)|^2 \leq Cn^{-3},$$

for every $n\in\mathbb{N}$. Note that, for the case that $\rho=2$, one obtains the same result immediately by using Cauchy-Schwarz inequality. As for $\mathbb{E}|J_9(t)|^2$, applying Remark 2.3 yields

$$\mathbb{E}|J_9(t)|^2 \leq C\int_{\kappa(n,t)}^{t}\mathbb{E}(1+|x_s^n|)^{\rho}(|\sigma_2^n(s,x_{\kappa(n,s)}^n)|^2+|\sigma_3^n(s,x_{\kappa(n,s)}^n)|^2)\,ds$$

which by applying Hölder's inequality gives

$$\mathbb{E}|J_9(t)|^2 \leq C\int_{\kappa(n,t)}^{t}(\mathbb{E}(1+|x_s^n|^{p_0}))^{\frac{\rho}{p_0}}\left(\mathbb{E}(|\sigma_2^n(s,x_{\kappa(n,s)}^n)|^{\frac{2p_0}{p_0-\rho}}+|\sigma_3^n(s,x_{\kappa(n,s)}^n)|^{\frac{2p_0}{p_0-\rho}})\right)^{\frac{p_0-\rho}{p_0}}\,ds,$$

for any $t\in[0,T]$. By Lemma 2.13, one obtains

$$\mathbb{E}|J_9(t)|^2 \leq Cn^{-3},$$

for every $n\in\mathbb{N}$. To estimate $\mathbb{E}|J_{10}(t)|^2$, one uses Young's inequality and Remark 2.3 to obtain

$$\mathbb{E}|J_{10}(t)|^2 \leq Cn^{-1}\int_{\kappa(n,t)}^{t}\mathbb{E}(1+|x_s^n|+|x_{\kappa(n,s)}^n|)^{3\rho}|x_s^n-x_{\kappa(n,s)}^n|^{2\alpha}\,ds$$

which implies due to Hölder's inequality

$$\mathbb{E}|J_{10}(t)|^2 \leq Cn^{-1}\int_{\kappa(n,t)}^{t}\left(\mathbb{E}(1+|x_{\kappa(n,s)}^n|)^{p_0}\right)^{\frac{3\rho}{p_0}}\left(\mathbb{E}|x_s^n-x_{\kappa(n,s)}^n|^{\frac{2\alpha p_0}{p_0-3\rho}}\right)^{\frac{p_0-3\rho}{p_0}}\,ds,$$

for any $t\in[0,T]$. Lemma 2.15 is used to obtain

$$\mathbb{E}|J_{10}(t)|^2 \leq Cn^{-(2+\alpha)},$$

for every $n\in\mathbb{N}$. Finally for $\mathbb{E}|J_{12}(t)|^2$, applying Young's inequality, Cauchy-Schwarz inequality and Remark 2.3 yield

$$\mathbb{E}|J_{12}(t)|^2 \leq Cn^{-1}\int_{\kappa(n,t)}^{t}\mathbb{E}(1+|x_s^n|)+|x_{\kappa(n,s)}^n|)^{2\rho}|\sigma_M^n(s,x_{\kappa(n,s)}^n)|^2\,ds$$

$$+ Cn^{-1}\int_{\kappa(n,t)}^{t} \mathbb{E}(1+|x_s^n|)^{\rho-2}|\tilde{\sigma}^n(x^n_{\kappa(n,s)})|^2|\sigma^n_M(s,x^n_{\kappa(n,s)})|^2\,ds,$$

which implies due to Hölder's inequality

$$\begin{aligned}
&\mathbb{E}|J_{12}(t)|^2\\
&\leq Cn^{-1}\int_{\kappa(n,t)}^{t}\Big(\mathbb{E}(1+|x_s^n|^{p_0}+|x^n_{\kappa(n,s)}|^{p_0})\Big)^{\frac{2\rho}{p_0}}\Big(\mathbb{E}|\sigma^n_M(s,x^n_{\kappa(n,s)})|^{\frac{2p_0}{p_0-2\rho}}\Big)^{\frac{p_0-2\rho}{p_0}}\,ds\\
&\quad + Cn^{-1}\int_{\kappa(n,t)}^{t}(\mathbb{E}(1+|x_s^n|)^{p_0})^{\frac{\rho-2}{p_0}}\Big(\mathbb{E}|\tilde{\sigma}^n(x^n_{\kappa(n,s)})|^{\frac{2p_0}{p_0-\rho+2}}|\sigma^n_M(x^n_{\kappa(n,s)})|^{\frac{2p_0}{p_0-\rho+2}}\Big)^{\frac{p_0-\rho+2}{p_0}}\,ds,
\end{aligned}$$

for any $t\in[0,T]$. Applying Lemma 2.11 and Lemma 2.13 to the first term and Cauchy-Schwarz inequality to the second term give

$$\mathbb{E}|J_{12}(t)|^2\leq Cn^{-3}+Cn^{-1}\int_{\kappa(n,t)}^{t}\Big(\mathbb{E}|\tilde{\sigma}^n(x^n_{\kappa(n,s)})|^{\frac{4p_0}{p_0-\rho+2}}\mathbb{E}|\sigma^n_M(x^n_{\kappa(n,s)})|^{\frac{4p_0}{p_0-\rho+2}}\Big)^{\frac{p_0-\rho+2}{2p_0}}\,ds,$$

which by using Lemma 2.13 yields the desired result, i.e.

$$\mathbb{E}|J_{12}(t)|^2\leq Cn^{-3},$$

for every $n\in\mathbb{N}$. Therefore, one obtains, for any $n\in\mathbb{N}$, $\alpha\in(0,1]$ and $p_0\geq 10\rho+2$,

$$\sup_{0\leq t\leq T}\mathbb{E}|\sigma(x_t^n)-\sigma(x^n_{\kappa(n,t)})-\sigma^n_M(t,x^n_{\kappa(n,t)})|^2\leq Cn^{-(2+\alpha)}+Cn^{-3}+Cn^{-5}\leq Cn^{-(2+\alpha)}.$$

□

Lemma 2.18. *Assume **A-1** to **A-5** hold and $p_0\geq 2(5\rho+1)$. Then, there exists a constant $C>0$, such that for any $n\in\mathbb{N}$,*

$$\sup_{0\leq t\leq T}\mathbb{E}|b(x_t^n)-b(x^n_{\kappa(n,t)})-b_1^n(t,x^n_{\kappa(n,t)})-b_2^n(t,x^n_{\kappa(n,t)})|^2\leq Cn^{-2}.$$

Proof. For every $k=1,\dots,d$, applying Itô's formula to $b^{(k)}(x_t^n)-b^{(k)}(x^n_{\kappa(n,t)})$ gives, almost surely,

$$\begin{aligned}
&b^{(k)}(x_t^n)-b^{(k)}(x^n_{\kappa(n,t)})\\
&=\sum_{i=1}^{d}\int_{\kappa(n,t)}^{t}\frac{\partial b^{(k)}(x_s^n)}{\partial x^{(i)}}\tilde{b}^{n,(i)}(s,x^n_{\kappa(n,s)})\,ds+\sum_{i=1}^{d}\sum_{j=1}^{m}\int_{\kappa(n,t)}^{t}\frac{\partial b^{(k)}(x_s^n)}{\partial x^{(i)}}\tilde{\sigma}^{n,(i,j)}(s,x^n_{\kappa(n,s)})\,dw_s^j\\
&\quad+\frac{1}{2}\sum_{i,l=1}^{d}\sum_{j=1}^{m}\int_{\kappa(n,t)}^{t}\frac{\partial^2 b^{(k)}(x_s^n)}{\partial x^{(i)}\partial x^{(l)}}\tilde{\sigma}^{n,(i,j)}(s,x^n_{\kappa(n,s)})\tilde{\sigma}^{n,(l,j)}(s,x^n_{\kappa(n,s)})\,ds\\
&=\sum_{i=1}^{9}I_i(t),
\end{aligned}\tag{2.12}$$

where

$$\begin{aligned}
I_1(t)&=\sum_{i=1}^{d}\int_{\kappa(n,t)}^{t}\left(\frac{\partial b^{(k)}(x_s^n)}{\partial x^{(i)}}-\frac{\partial b^{(k)}(x^n_{\kappa(n,s)})}{\partial x^{(i)}}\right)b^{n,(i)}(x^n_{\kappa(n,s)})\,ds,\\
I_2(t)&=\sum_{i=1}^{d}\int_{\kappa(n,t)}^{t}\frac{\partial b^{(k)}(x^n_{\kappa(n,s)})}{\partial x^{(i)}}b^{n,(i)}(x^n_{\kappa(n,s)})\,ds,
\end{aligned}$$

$$I_3(t) = \sum_{i=1}^{d} \int_{\kappa(n,t)}^{t} \frac{\partial b^{(k)}(x_s^n)}{\partial x^{(i)}} (b_1^{n,(i)}(s, x_{\kappa(n,s)}^n) + b_2^{n,(i)}(s, x_{\kappa(n,s)}^n))\, ds,$$

$$I_4(t) = \sum_{i=1}^{d} \sum_{j=1}^{m} \int_{\kappa(n,t)}^{t} \left(\frac{\partial b^{(k)}(x_s^n)}{\partial x^{(i)}} - \frac{\partial b^{(k)}(x_{\kappa(n,s)}^n)}{\partial x^{(i)}} \right) \sigma^{n,(i,j)}(x_{\kappa(n,s)}^n)\, dw_s^j,$$

$$I_5(t) = \sum_{i=1}^{d} \sum_{j=1}^{m} \int_{\kappa(n,t)}^{t} \frac{\partial b^{(k)}(x_{\kappa(n,s)}^n)}{\partial x^{(i)}} \sigma^{n,(i,j)}(x_{\kappa(n,s)}^n)\, dw_s^j,$$

$$I_6(t) = \sum_{i=1}^{d} \sum_{j=1}^{m} \int_{\kappa(n,t)}^{t} \frac{\partial b^{(k)}(x_s^n)}{\partial x^{(i)}} \sigma_M^{n,(i,j)}(s, x_{\kappa(n,s)}^n)\, dw_s^j,$$

$$I_7(t) = \frac{1}{2} \sum_{i,l=1}^{d} \sum_{j=1}^{m} \int_{\kappa(n,t)}^{t} \left(\frac{\partial^2 b^{(k)}(x_s^n)}{\partial x^{(i)} \partial x^{(l)}} - \frac{\partial^2 b^{(k)}(x_{\kappa(n,s)}^n)}{\partial x^{(i)} \partial x^{(l)}} \right) \sigma^{n,(i,j)}(x_{\kappa(n,s)}^n) \sigma^{n,(l,j)}(x_{\kappa(n,s)}^n)\, ds,$$

$$I_8(t) = \frac{1}{2} \sum_{i,l=1}^{d} \sum_{j=1}^{m} \int_{\kappa(n,t)}^{t} \frac{\partial^2 b^{(k)}(x_{\kappa(n,s)}^n)}{\partial x^{(i)} \partial x^{(l)}} \sigma^{n,(i,j)}(x_{\kappa(n,s)}^n) \sigma^{n,(l,j)}(x_{\kappa(n,s)}^n)\, ds,$$

$$\begin{aligned} I_9(t) = \frac{1}{2} \sum_{i,l=1}^{d} \sum_{j=1}^{m} \int_{\kappa(n,t)}^{t} \frac{\partial^2 b^{(k)}(x_s^n)}{\partial x^{(i)} \partial x^{(l)}} (&\sigma^{n,(i,j)}(x_{\kappa(n,s)}^n) \sigma_M^{n,(l,j)}(s, x_{\kappa(n,s)}^n) \\ &+ \sigma_M^{n,(i,j)}(s, x_{\kappa(n,s)}^n) \tilde{\sigma}^{n,(l,j)}(s, x_{\kappa(n,s)}^n))\, ds. \end{aligned}$$

Note that

$$\begin{aligned} &\mathbb{E}|I_2(t) + I_8(t) - b_1^{n,(k)}(t, x_{\kappa(n,t)}^n)|^2 \\ &\leq C \sum_{i,l=1}^{d} \sum_{j=1}^{m} \mathbb{E} \left| -\frac{n^{-3/2}|x_{\kappa(n,t)}^n|^{3\rho}}{(1 + n^{-3/2}|x_{\kappa(n,t)}^n|^{3\rho})^2} \int_{\kappa(n,t)}^{t} \frac{\partial^2 b^{(k)}(x_{\kappa(n,s)}^n)}{\partial x^{(i)} \partial x^{(l)}} \sigma^{(i,j)}(x_{\kappa(n,s)}^n) \sigma^{(l,j)}(x_{\kappa(n,s)}^n)\, ds \right|^2, \end{aligned}$$

which by applying Remark 2.3 and Lemma 2.11 yields

$$\mathbb{E}|I_2(t) + I_8(t) - b_1^{n,(k)}(t, x_{\kappa(n,t)}^n)|^2 \leq C n^{-5} \mathbb{E}||x_{\kappa(n,t)}^n|^{3\rho}(1 + |x_{\kappa(n,t)}^n|^{2\rho+1})|^2 \leq C n^{-5}, \tag{2.13}$$

for $p_0 \geq 10\rho + 2$. Moreover, notice that

$$I_5(t) = b_2^{n,(k)}(t, x_{\kappa(n,t)}^n). \tag{2.14}$$

Then, one obtains the following

$$\begin{aligned} &\mathbb{E}|b^{(k)}(x_t^n) - b^{(k)}(x_{\kappa(n,t)}^n) - b_1^{n,(k)}(t, x_{\kappa(n,t)}^n) - b_2^{n,(k)}(t, x_{\kappa(n,t)}^n)|^2 \\ &\quad \leq 2\mathbb{E}|I_1(t) + I_3(t) + I_4(t) + I_6(t) + I_7(t) + I_9(t)|^2 + 2\mathbb{E}|I_2(t) + I_8(t) - b_1^{n,(k)}(t, x_{\kappa(n,t)}^n)|^2 \\ &\quad \leq C(\mathbb{E}|I_1(t)|^2 + \mathbb{E}|I_3(t)|^2 + \mathbb{E}|I_4(t)|^2 + \mathbb{E}|I_6(t)|^2 + \mathbb{E}|I_7(t)|^2 + \mathbb{E}|I_9(t)|^2) + C n^{-5}, \end{aligned}$$

for any $t \in [0, T]$. To estimate $\mathbb{E}|I_1(t)|^2$, applying Cauchy-Schwarz inequality and Remark 2.3 yield

$$\mathbb{E}|I_1(t)|^2 \leq C n^{-1} \int_{\kappa(n,t)}^{t} \mathbb{E}(1 + |x_s^n| + |x_{\kappa(n,s)}^n|)^{2\rho-2}(1 + |x_{\kappa(n,s)}^n|)^{2\rho+2}|x_s^n - x_{\kappa(n,s)}^n|^2\, ds,$$

which further implies due to Young's inequality and Hölder's inequality

$$\mathbb{E}|I_1(t)|^2 \leq C n^{-1} \int_{\kappa(n,t)}^{t} \left(\mathbb{E}(1 + |x_s^n|^{p_0} + |x_{\kappa(n,s)}^n|^{p_0}) \right)^{\frac{4\rho}{p_0}} \left(\mathbb{E}|x_s^n - x_{\kappa(n,s)}^n|^{\frac{2p_0}{p_0 - 4\rho}} \right)^{\frac{p_0 - 4\rho}{p_0}}\, ds,$$

for any $t \in [0, T]$. By Lemma 2.15, one obtains

$$\mathbb{E}|I_1(t)|^2 \leq Cn^{-3},$$

for any $n \in \mathbb{N}$. As for $\mathbb{E}|I_3(t)|^2$, applying Cauchy-Schwarz inequality and Remark 2.3 give

$$\mathbb{E}|I_3(t)|^2 \leq Cn^{-1} \int_{\kappa(n,t)}^{t} \mathbb{E}(1+|x_s^n|)^{2\rho}(|b_1^n(s,x_{\kappa(n,s)}^n)|^2 + |b_2^n(s,x_{\kappa(n,s)}^n)|^2)\,ds,$$

then one writes by using Hölder's inequality that

$$\begin{aligned}\mathbb{E}|I_3(t)|^2 \leq Cn^{-1} \int_{\kappa(n,t)}^{t} \Big((\mathbb{E}(1+|x_s^n|^{p_0}))^{\frac{2\rho}{p_0}} \Big(\mathbb{E}|b_1^n(s,x_{\kappa(n,s)}^n)|^{\frac{2p_0}{p_0-2\rho}} \\ + \mathbb{E}|b_2^n(s,x_{\kappa(n,s)}^n)|^{\frac{2p_0}{p_0-2\rho}} \Big)^{\frac{p_0-2\rho}{p_0}} \Big)\,ds,\end{aligned}$$

for any $t \in [0, T]$. Applying Lemma 2.13 yields

$$\mathbb{E}|I_3(t)|^2 \leq Cn^{-3},$$

for any $n \in \mathbb{N}$. To estimate $\mathbb{E}|I_4(t)|^2$, one uses Cauchy-Schwarz inequality, Remark 2.3 and Young's inequality to obtain

$$\begin{aligned}\mathbb{E}|I_4(t)|^2 &\leq C \int_{\kappa(n,t)}^{t} \mathbb{E}(1+|x_s^n|+|x_{\kappa(n,s)}^n|)^{2\rho-2}(1+|x_{\kappa(n,s)}^n|)^{\rho+2}|x_s^n - x_{\kappa(n,s)}^n|^2\,ds \\ &\leq C \int_{\kappa(n,t)}^{t} \mathbb{E}(1+|x_s^n|+|x_{\kappa(n,s)}^n|)^{3\rho}|x_s^n - x_{\kappa(n,s)}^n|^2\,ds,\end{aligned}$$

which implies due to Young's inequality and Hölder's inequality

$$\mathbb{E}|I_4(t)|^2 \leq C \int_{\kappa(n,t)}^{t} \Big(\mathbb{E}(1+|x_s^n|^{p_0}+|x_{\kappa(n,s)}^n|^{p_0}) \Big)^{\frac{3\rho}{p_0}} \Big(\mathbb{E}|x_s^n - x_{\kappa(n,s)}^n|^{\frac{2p_0}{p_0-3\rho}} \Big)^{\frac{p_0-3\rho}{p_0}}\,ds,$$

for any $t \in [0, T]$. One applies Lemma 2.15 to obtain

$$\mathbb{E}|I_4(t)|^2 \leq Cn^{-2}, \tag{2.15}$$

for any $n \in \mathbb{N}$. As for $\mathbb{E}|I_6(t)|^2$, it can be written as

$$\mathbb{E}|I_6(t)|^2 \leq C \int_{\kappa(n,t)}^{t} \mathbb{E}(1+|x_s^n|)^{2\rho}|\sigma_M^n(s,x_{\kappa(n,s)}^n)|^2\,ds,$$

which by using Hölder's inequality yields

$$\mathbb{E}|I_6(t)|^2 \leq C \int_{\kappa(n,t)}^{t} (\mathbb{E}(1+|x_s^n|^{p_0}))^{\frac{2\rho}{p_0}} \Big(\mathbb{E}|\sigma_M^n(s,x_{\kappa(n,s)}^n)|^{\frac{2p_0}{p_0-2\rho}} \Big)^{\frac{p_0-2\rho}{p_0}}\,ds,$$

for any $t \in [0, T]$. By using Lemma 2.11 and Lemma 2.13, one obtains

$$\mathbb{E}|I_6(t)|^2 \leq Cn^{-2}, \tag{2.16}$$

for any $n \in \mathbb{N}$. In order to estimate $\mathbb{E}|I_7(t)|^2$, one uses Cauchy-Schwarz inequality and Remark 2.3 to obtain

$$\mathbb{E}|I_7(t)|^2 \leq Cn^{-1} \int_{\kappa(n,t)}^{t} \mathbb{E}(1+|x_s^n|+|x_{\kappa(n,s)}^n|)^{2\rho-4}(1+|x_{\kappa(n,s)}^n|)^{2\rho+4}|x_s^n - x_{\kappa(n,s)}^n|^2\,ds$$

$$\leq Cn^{-1}\int_{\kappa(n,t)}^{t}\mathbb{E}(1+|x_s^n|+|x_{\kappa(n,s)}^n|)^{4\rho}|x_s^n-x_{\kappa(n,s)}^n|^2\,ds,$$

which by applying Young's inequality and Hölder's inequality yields

$$\mathbb{E}|I_7(t)|^2\leq Cn^{-1}\int_{\kappa(n,t)}^{t}\left(\mathbb{E}(1+|x_s^n|+|x_{\kappa(n,s)}^n|)^{p_0}\right)^{\frac{4\rho}{p_0}}\left(\mathbb{E}|x_s^n-x_{\kappa(n,s)}^n|^{\frac{2p_0}{p_0-4\rho}}\right)^{\frac{p_0-4\rho}{p_0}}\,ds,$$

for any $t\in[0,T]$. Then applying Lemma 2.11 and Lemma 2.15, one obtains

$$\mathbb{E}|I_7(t)|^2\leq Cn^{-3},$$

for any $n\in\mathbb{N}$. Finally for $\mathbb{E}|I_9(t)|^2$, one writes

$$\begin{aligned}\mathbb{E}|I_9(t)|^2\leq{}& Cn^{-1}\int_{\kappa(n,t)}^{t}\mathbb{E}(1+|x_s^n|)^{2\rho-2}(1+|x_{\kappa(n,s)}^n|)^{\rho+2}|\sigma_M^n(s,x_{\kappa(n,s)}^n)|^2\,ds\\&+Cn^{-1}\int_{\kappa(n,t)}^{t}\mathbb{E}(1+|x_s^n|)^{2\rho-2}|\sigma_M^n(s,x_{\kappa(n,s)}^n)|^2|\tilde{\sigma}^n(s,x_{\kappa(n,s)}^n)|^2\,ds,\end{aligned}$$

which implies due to Young's inequality and Hölder's inequality

$$\begin{aligned}\mathbb{E}|I_9(t)|^2\leq{}& Cn^{-1}\int_{\kappa(n,t)}^{t}\left(\mathbb{E}(1+|x_s^n|+|x_{\kappa(n,s)}^n|)^{p_0}\right)^{\frac{3\rho}{p_0}}\left(\mathbb{E}|\sigma_M^n(s,x_{\kappa(n,s)}^n|^{\frac{2p_0}{p_0-3\rho}}\right)^{\frac{p_0-3\rho}{p_0}}\,ds\\&+Cn^{-1}\int_{\kappa(n,t)}^{t}(\mathbb{E}(1+|x_s^n|)^{p_0})^{\frac{2\rho-2}{p_0}}\\&\qquad\times\left(\mathbb{E}|\tilde{\sigma}^n(s,x_{\kappa(n,s)}^n)|^{\frac{2p_0}{p_0-2\rho+2}}|\sigma_M^n(s,x_{\kappa(n,s)}^n)|^{\frac{2p_0}{p_0-2\rho+2}}\right)^{\frac{p_0-2\rho+2}{p_0}}\,ds\end{aligned}$$

for any $t\in[0,T]$. One can then apply Lemma 2.11 and Lemma 2.13 to the first term, and apply Cauchy-Schwarz inequality to the second term to obtain

$$\mathbb{E}|I_9(t)|^2\leq Cn^{-3}+Cn^{-1}\int_{\kappa(n,t)}^{t}\left(\mathbb{E}|\tilde{\sigma}^n(s,x_{\kappa(n,s)}^n)|^{\frac{4p_0}{p_0-2\rho+2}}\mathbb{E}|\sigma_M^n(s,x_{\kappa(n,s)}^n)|^{\frac{4p_0}{p_0-2\rho+2}}\right)^{\frac{p_0-2\rho+2}{2p_0}}\,ds$$

which, by using Lemma 2.13, implies

$$\mathbb{E}|I_9(t)|^2\leq Cn^{-3},$$

for any $n\in\mathbb{N}$ and $t\in[0,T]$. Therefore,

$$\sup_{0\leq t\leq T}\mathbb{E}|b(x_t^n)-b(x_{\kappa(n,t)}^n)-b_1^n(t,x_{\kappa(n,t)}^n)-b_2^n(t,x_{\kappa(n,t)}^n)|^2\leq Cn^{-2}+Cn^{-5}\leq Cn^{-2},$$

for any $n\in\mathbb{N}$, and the proof is complete. □

Denote by $e_t^n:=x_t-x_t^n$ for any $t\in[0,T]$, and define the stopping times as follows: for $R>0$,

$$.\tau_R:=\inf\{t\geq 0:|x_t|\geq R\},\quad \tau'_{n,R}:=\inf\{t\geq 0:|x_t^n|\geq R\},\quad \nu_{n,R}:=\tau_R\wedge\tau'_{n,R}.\tag{2.17}$$

Lemma 2.19. *Assume **A-1** to **A-5** hold and $p_0\geq 2(5\rho+1)$. Then, there exists a constant $C>0$ such that for any $s\in[0,T]$, the following inequality holds*

$$\mathbb{P}(s>\nu_{n,R})\leq CR^{-2},$$

where $\nu_{n,R}$ is the stopping time defined in (2.17).

Proof. By applying Markov inequality, one obtains

$$\begin{aligned}\mathbb{P}(s > \nu_{n,R}) &\le \mathbb{P}\left(\sup_{u\le s}|x_u| > R\right) + \mathbb{P}\left(\sup_{u\le s}|x_u^n| > R\right)\\ &\le R^{-2}\mathbb{E}\left(\sup_{u\le s}|x_u|^2\right) + R^{-2}\mathbb{E}\left(\sup_{u\le s}|x_u^n|^2\right)\\ &\le CR^{-2}.\end{aligned}$$

Note that the last inequality holds since by Lemma 2.7 and Lemma 2.11, we have shown that the p_0-th moment of x_t and x_t^n are bounded uniformly in time, i.e. $\sup_{0\le t\le T}\mathbb{E}|x_t|^{p_0} \le C$ and $\sup_{0\le t\le T}\mathbb{E}|x_t^n|^{p_0} \le C$ for all $n \in \mathbb{N}$ and $p_0 \ge 4$. Then, one can obtain the uniform $\mathscr{L}^2$ bound by using Lemma 5 in [45], which originally appeared in [22]. □

Lemma 2.20. *Assume **A-1** to **A-5** hold and $p_0 \ge 2(5\rho+1)$. Then, there exists a constant $C > 0$, which is independent of R, such that for any $n \in \mathbb{N}$ and $t \in [0,T]$,*

$$\begin{aligned}&\mathbb{E}\int_0^{t\wedge\nu_{n,R}} e_s^n(b(x_s^n) - b(x_{\kappa(n,s)}^n) - b_1^n(s, x_{\kappa(n,s)}^n) - b_2^n(s, x_{\kappa(n,s)}^n))\,ds\\ &\qquad\qquad \le C\int_0^t \sup_{0\le r\le s}\mathbb{E}|e_{r\wedge\nu_{n,R}}^n|^2\,ds + Cn^{-\frac{5+\alpha}{2}} + CR^{-\frac{2}{5}}n^{-2},\end{aligned}$$

where $\nu_{n,R}$ is the stopping time defined in (2.17).

Proof. First, for any $k = 1, \ldots, d$, applying Itô's formula to $b^{(k)}(x_t^n) - b^{(k)}(x_{\kappa(n,t)}^n)$ gives (2.12). Then, by (2.13) and (2.14), one obtains

$$\mathbb{E}\int_0^{t\wedge\nu_{n,R}} e_s^{n,(k)}(b^{(k)}(x_s^n) - b^{(k)}(x_{\kappa(n,s)}^n) - b_1^{n,(k)}(s, x_{\kappa(n,s)}^n) - b_2^{n,(k)}(s, x_{\kappa(n,s)}^n))\,ds \le \sum_{i=1}^7 T_i(t) + T_8,$$

where

$$\begin{aligned}
T_1(t) &= \mathbb{E}\int_0^{t\wedge\nu_{n,R}} e_s^{n,(k)}\sum_{i=1}^d\int_{\kappa(n,s)}^s\left(\frac{\partial b^{(k)}(x_r^n)}{\partial x^{(i)}} - \frac{\partial b^{(k)}(x_{\kappa(n,r)}^n)}{\partial x^{(i)}}\right)b^{n,(i)}(x_{\kappa(n,r)}^n)\,dr\,ds,\\
T_2(t) &= \mathbb{E}\int_0^{t\wedge\nu_{n,R}} e_s^{n,(k)}\sum_{i=1}^d\int_{\kappa(n,s)}^s\frac{\partial b^{(k)}(x_r^n)}{\partial x^{(i)}}(b_1^{n,(i)}(r, x_{\kappa(n,r)}^n) + b_2^{n,(i)}(r, x_{\kappa(n,r)}^n))\,dr\,ds,\\
T_3(t) &= \mathbb{E}\int_0^{t\wedge\nu_{n,R}} e_s^{n,(k)}\sum_{i=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s\left(\frac{\partial b^{(k)}(x_r^n)}{\partial x^{(i)}} - \frac{\partial b^{(k)}(x_{\kappa(n,r)}^n)}{\partial x^{(i)}}\right)\sigma^{n,(i,j)}(x_{\kappa(n,r)}^n)\,dw_r^j\,ds,\\
T_4(t) &= \mathbb{E}\int_0^{t\wedge\nu_{n,R}} e_s^{n,(k)}\sum_{i=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s\frac{\partial b^{(k)}(x_r^n)}{\partial x^{(i)}}\sigma_M^{n,(i,j)}(r, x_{\kappa(n,r)}^n)\,dw_r^j\,ds,\\
T_5(t) &= \frac{1}{2}\mathbb{E}\int_0^{t\wedge\nu_{n,R}} e_s^{n,(k)}\sum_{i,l=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s\left(\frac{\partial^2 b^{(k)}(x_r^n)}{\partial x^{(i)}\partial x^{(l)}} - \frac{\partial^2 b^{(k)}(x_{\kappa(n,r)}^n)}{\partial x^{(i)}\partial x^{(l)}}\right)\\
&\qquad\qquad\qquad\qquad \times\sigma^{n,(i,j)}(x_{\kappa(n,r)}^n)\sigma^{n,(l,j)}(x_{\kappa(n,r)}^n)\,dr\,ds,\\
T_6(t) &= \frac{1}{2}\mathbb{E}\int_0^{t\wedge\nu_{n,R}} e_s^{n,(k)}\sum_{i,l=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s\frac{\partial^2 b^{(k)}(x_r^n)}{\partial x^{(i)}\partial x^{(l)}}(\sigma^{n,(i,j)}(x_{\kappa(n,r)}^n)\sigma_M^{n,(l,j)}(r, x_{\kappa(n,r)}^n)\\
&\qquad\qquad\qquad\qquad + \sigma_M^{n,(i,j)}(r, x_{\kappa(n,r)}^n)\tilde{\sigma}^{n,(l,j)}(r, x_{\kappa(n,r)}^n))\,dr\,ds,\\
T_7(t) &= C\int_0^t\sup_{0\le r\le s}\mathbb{E}|e_{r\wedge\nu_{n,R}}^n|^2\,ds,\\
T_8(t) &= Cn^{-5}.
\end{aligned}$$

To estimate $T_1(t)$, one applies Young's inequality and Remark 2.3 to obtain

$$T_1(t) \le C\int_0^t \sup_{0\le r\le s} \mathbb{E}|e^n_{r\wedge\nu_{n,R}}|^2\,ds + Cn^{-1}\int_0^t\int_{\kappa(n,s)}^s \mathbb{E}(1+|x^n_r|+|x^n_{\kappa(n,r)}|)^{4\rho}|x^n_r-x^n_{\kappa(n,r)}|^2\,dr\,ds,$$

which by using Hölder's inequality implies

$$T_1(t) \le C\int_0^t \sup_{0\le r\le s} \mathbb{E}|e^n_{r\wedge\nu_{n,R}}|^2\,ds + Cn^{-1}\int_0^t\int_{\kappa(n,s)}^s \Big(\mathbb{E}(1+|x^n_r|^{p_0}+|x^n_{\kappa(n,r)}|^{p_0})\Big)^{\frac{4\rho}{p_0}}\Big(\mathbb{E}|x^n_r-x^n_{\kappa(n,r)}|^{\frac{2p_0}{p_0-4\rho}}\Big)^{\frac{p_0-4\rho}{p_0}}\,dr\,ds.$$

Thus, by Lemma 2.15 and Lemma 2.11, one obtains

$$T_1(t) \le C\int_0^t \sup_{0\le r\le s} \mathbb{E}|e^n_{r\wedge\nu_{n,R}}|^2\,ds + Cn^{-3},$$

for any $n\in\mathbb{N}$. For $T_2(t)$, $T_5(t)$ and $T_6(t)$, the same results can be obtained by the direct application of Cauchy-Schwarz inequality combining with previous Lemmas and Remarks. The rest of the proof will mainly focus on obtaining estimates for $T_3(t)$ and $T_4(t)$.

For any $r\in[0,T]$, $i,k=1,\dots,d$ and $j=1,\dots,m$, denote by

$$\mathbb{T}^{(i,j,k)}_r := \left(\frac{\partial b^{(k)}(x^n_r)}{\partial x^{(i)}} - \frac{\partial b^{(k)}(x^n_{\kappa(n,r)})}{\partial x^{(i)}}\right)\sigma^{n,(i,j)}(x^n_{\kappa(n,r)}) + \frac{\partial b^{(k)}(x^n_r)}{\partial x^{(i)}}\sigma^{n,(i,j)}_M(r,x^n_{\kappa(n,r)}).$$

Then, applying Remark 2.3 and Hölder's inequality yields

$$\begin{aligned}\mathbb{E}|\mathbb{T}^{(i,j,k)}_r|^p &= \mathbb{E}\left|\left(\frac{\partial b^{(k)}(x^n_r)}{\partial x^{(i)}} - \frac{\partial b^{(k)}(x^n_{\kappa(n,r)})}{\partial x^{(i)}}\right)\sigma^{n,(i,j)}(x^n_{\kappa(n,r)}) + \frac{\partial b^{(k)}(x^n_r)}{\partial x^{(i)}}\sigma^{n,(i,j)}_M(r,x^n_{\kappa(n,r)})\right|^p \\ &\le C\Big(\mathbb{E}(1+|x^n_r|+|x^n_{\kappa(n,r)}|)^{p_0}\Big)^{\frac{3\rho p}{2p_0}}\Big(\mathbb{E}|x^n_r-x^n_{\kappa(n,r)}|^{\frac{2pp_0}{2p_0-3\rho p}}\Big)^{\frac{2p_0-3\rho p}{2p_0}} \\ &\qquad + C\left(\mathbb{E}(1+|x^n_r|)^{p_0}\right)^{\frac{\rho p}{p_0}}\left(\mathbb{E}|\sigma^{n,(i,j)}_M(r,x^n_{\kappa(n,r)})|^{\frac{pp_0}{p_0-\rho p}}\right)^{\frac{p_0-\rho p}{p_0}},\end{aligned}$$

which, by using Lemma 2.15 and Lemma 2.13, implies

$$\sup_{r\le T}\mathbb{E}|\mathbb{T}^{(i,j,k)}_r|^p \le Cn^{-\frac{p}{2}}, \tag{2.18}$$

for $p\le\frac{2p_0}{7\rho+2}$. Due to (2.15) and (2.16) in the proof of Lemma 2.18, one can also obtain the following estimate

$$\mathbb{E}\left|\sum_{i=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s \mathbb{T}^{(i,j,k)}_r\,dw^j_r\right|^2 \le 2\mathbb{E}|I_4(t)|^2+2\mathbb{E}|I_6(t)|^2 \le Cn^{-2}. \tag{2.19}$$

Then, one writes

$$\begin{aligned}T_3(t)+T_4(t) :=& \mathbb{E}\int_0^{t\wedge\nu_{n,R}} e^{n,(k)}_s\sum_{i=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s \mathbb{T}^{(i,j,k)}_r\,dw^j_r\,ds \\ =& \mathbb{E}\int_0^{t\wedge\nu_{n,R}} (e^{n,(k)}_s - e^{n,(k)}_{\kappa(n,s)})\sum_{i=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s \mathbb{T}^{(i,j,k)}_r\,dw^j_r\,ds\end{aligned}$$

$$+\mathbb{E}\int_0^{t\wedge\nu_{n,R}} e^{n,(k)}_{\kappa(n,s)}\sum_{i=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s \mathbb{T}_r^{(i,j,k)}\,dw_r^j\,ds.$$

Note that the second term above is not zero. However, by using Lemma 2.19, one obtains

$$\begin{aligned}
&\mathbb{E}\int_0^{t\wedge\nu_{n,R}} e^{n,(k)}_{\kappa(n,s)}\sum_{i=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s \mathbb{T}_r^{(i,j,k)}\,dw_r^j\,ds\\
&=\mathbb{E}\int_0^t \mathbb{1}_{\{s\le\nu_{n,R}\}} e^{n,(k)}_{\kappa(n,s)\wedge\nu_{n,R}}\sum_{i=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s \mathbb{T}_r^{(i,j,k)}\,dw_r^j\,ds\\
&=\mathbb{E}\int_0^t e^{n,(k)}_{\kappa(n,s)\wedge\nu_{n,R}}\sum_{i=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s \mathbb{T}_r^{(i,j,k)}\,dw_r^j\,ds\\
&\qquad-\mathbb{E}\int_0^t \mathbb{1}_{\{s>\nu_{n,R}\}} e^{n,(k)}_{\kappa(n,s)\wedge\nu_{n,R}}\sum_{i=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s \mathbb{T}_r^{(i,j,k)}\,dw_r^j\,ds,
\end{aligned}$$

where the first term is zero since $\kappa(n,s)\wedge\nu_{n,R}$ is $\mathscr{F}_{\kappa(n,s)}$-measurable. Then, applying Young's inequality, Hölder's inequality to the second term yield

$$\begin{aligned}
&\mathbb{E}\int_0^{t\wedge\nu_{n,R}} e^{n,(k)}_{\kappa(n,s)}\sum_{i=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s \mathbb{T}_r^{(i,j,k)}\,dw_r^j\,ds\\
&\le C\int_0^t\sup_{0\le r\le s}\mathbb{E}|e^n_{r\wedge\nu_{n,R}}|^2\,ds+C\int_0^t\left(\mathbb{P}\left(s>\nu_{n,R}\right)\right)^{\frac15}\left(\mathbb{E}\left|\sum_{i=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s \mathbb{T}_r^{(i,j,k)}\,dw_r^j\right|^{\frac52}\right)^{\frac45}ds\\
&\le C\int_0^t\sup_{0\le r\le s}\mathbb{E}|e^n_{r\wedge\nu_{n,R}}|^2\,ds+CR^{-\frac25}n^{-2},
\end{aligned}$$

where the last inequality holds due to (2.18). One may notice that (2.18) holds only when $\frac52\le\frac{2p_0}{7\rho+2}$, which implies $p_0\ge\frac54(7\rho+2)$. However, as $\frac54(7\rho+2)\le 2(5\rho+1)$ for all $\rho\ge2$, by assuming $p_0\ge 2(5\rho+1)$, (2.18) holds automatically for $p=\frac52$. Furthermore, $T_3(t)+T_4(t)$ can be expressed as

$$\begin{aligned}
T_3(t)+T_4(t)=&\,\mathbb{E}\int_0^{t\wedge\nu_{n,R}}\int_{\kappa(n,s)}^s \bar b^{n,(k)}(r,x^n_{\kappa(n,r)})\,dr\sum_{i=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s \mathbb{T}_r^{(i,j,k)}\,dw_r^j\,ds\\
&+\mathbb{E}\int_0^{t\wedge\nu_{n,R}}\sum_{v=1}^m\int_{\kappa(n,s)}^s \bar\sigma^{n,(k,v)}(r,x^n_{\kappa(n,r)})\,dw_r^v\sum_{i=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s \mathbb{T}_r^{(i,j,k)}\,dw_r^j\,ds\\
&+C\int_0^t\sup_{0\le r\le s}\mathbb{E}|e^n_{r\wedge\nu_{n,R}}|^2\,ds+CR^{-\frac25}n^{-2},
\end{aligned}$$

where

$$\bar b^{n,(k)}(t,x^n_{\kappa(n,t)})=b^{(k)}(x_t)-\tilde b^{n,(k)}(t,x^n_{\kappa(n,t)})$$

and

$$\bar\sigma^{n,(k,v)}(t,x^n_{\kappa(n,t)})=\sigma^{(k,v)}(x_t)-\tilde\sigma^{n,(k,v)}(t,x^n_{\kappa(n,t)}).$$

One observes that $T_3(t)+T_4(t)$ can be expanded as

$$\begin{aligned}
&T_3(t)+T_4(t)\\
&=\mathbb{E}\int_0^{t\wedge\nu_{n,R}}\int_{\kappa(n,s)}^s (b^{(k)}(x_r)-b^{(k)}(x_r^n))\,dr\sum_{i=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s \mathbb{T}_r^{(i,j,k)}\,dw_r^j\,ds
\end{aligned}$$

$$
\begin{aligned}
&+\mathbb{E}\int_0^{t\wedge\nu_{n,R}}\int_{\kappa(n,s)}^s (b^{(k)}(x_r^n)-b^{(k)}(x_{\kappa(n,r)}^n)-b_1^{n,(k)}(r,x_{\kappa(n,r)}^n)-b_2^{n,(k)}(r,x_{\kappa(n,r)}^n))\,dr\\
&\qquad\qquad\times\sum_{i=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s \mathbb{T}_r^{(i,j,k)}\,dw_r^j\,ds\\
&+\mathbb{E}\int_0^{t\wedge\nu_{n,R}}\int_{\kappa(n,s)}^s (b^{(k)}(x_{\kappa(n,r)}^n)-b^{n,(k)}(x_{\kappa(n,r)}^n)\,dr\sum_{i=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s \mathbb{T}_r^{(i,j,k)}\,dw_r^j\,ds\\
&+\sum_{v=j=1}^m\sum_{i=1}^d\mathbb{E}\int_0^{t\wedge\nu_{n,R}}\int_{\kappa(n,s)}^s (\sigma^{(k,v)}(x_r)-\sigma^{(k,v)}(x_r^n))\mathbb{T}_r^{(i,j,k)}\,dr\,ds\\
&+\sum_{v=j=1}^m\sum_{i=1}^d\mathbb{E}\int_0^{t\wedge\nu_{n,R}}\int_{\kappa(n,s)}^s (\sigma^{(k,v)}(x_r^n)-\sigma^{(k,v)}(x_{\kappa(n,r)}^n)-\sigma_M^{n,(k,v)}(r,x_{\kappa(n,r)}^n))\mathbb{T}_r^{(i,j,k)}\,dr\,ds\\
&+\sum_{v=j=1}^m\sum_{i=1}^d\mathbb{E}\int_0^{t\wedge\nu_{n,R}}\int_{\kappa(n,s)}^s (\sigma^{(k,v)}(x_{\kappa(n,r)}^n)-\sigma^{n,(k,v)}(x_{\kappa(n,r)}^n))\mathbb{T}_r^{(i,j,k)}\,dr\,ds\\
&+C\int_0^t\sup_{0\le r\le s}\mathbb{E}|e_{r\wedge\nu_{n,R}}^n|^2\,ds+CR^{-\frac{2}{5}}n^{-2},
\end{aligned}
$$

which implies due to Remark 2.3, Young's inequality and Cauchy-Schwarz inequality

$$
\begin{aligned}
&T_3(t)+T_4(t)\\
&\le C\mathbb{E}\int_0^{t\wedge\nu_{n,R}}\left(\int_{\kappa(n,s)}^s(1+|x_r|+|x_r^n|)^{2\rho}\,dr\int_{\kappa(n,s)}^s|e_r^n|^2\,dr\right)^{\frac{1}{2}}\left|\sum_{i=1}^d\sum_{j=1}^m\int_{\kappa(n,s)}^s\mathbb{T}_r^{(i,j,k)}\,dw_r^j\right|ds\\
&\quad+C\int_0^t\left(\mathbb{E}\left|\int_{\kappa(n,s)}^s|b(x_r^n)-b(x_{\kappa(n,r)}^n)-b_1^n(r,x_{\kappa(n,r)}^n)-b_2^n(r,x_{\kappa(n,r)}^n)|\,dr\right|^2\right.\\
&\qquad\qquad\left.\times\sum_{i=1}^d\sum_{j=1}^m\mathbb{E}\left|\int_{\kappa(n,s)}^s\mathbb{T}_r^{(i,j,k)}\,dw_r^j\right|^2\right)^{1/2}ds\\
&\quad+C\int_0^t n\times n^{-1}\mathbb{E}\int_{\kappa(n,s)}^s|b(x_{\kappa(n,r)}^n)-b^n(x_{\kappa(n,r)}^n)|^2\,dr\,ds\\
&\quad+Cn^{-1}\sum_{i=1}^d\sum_{j=1}^m\int_0^t\mathbb{E}\left|\int_{\kappa(n,s)}^s\mathbb{T}_r^{(i,j,k)}\,dw_r^j\right|^2ds\\
&\quad+C\sum_{i=1}^d\sum_{j=1}^m\mathbb{E}\int_0^{t\wedge\nu_{n,R}}\int_{\kappa(n,s)}^s(1+|x_r|+|x_r^n|)^{\frac{\rho}{2}}|e_r^n||\mathbb{T}_r^{(i,j,k)}|\,dr\,ds\\
&\quad+C\sum_{i=1}^d\sum_{j=1}^m\int_0^t\int_{\kappa(n,s)}^s\sqrt{\mathbb{E}|\sigma(x_r^n)-\sigma(x_{\kappa(n,r)}^n)-\sigma_M^n(r,x_{\kappa(n,r)}^n)|^2\mathbb{E}|\mathbb{T}_r^{(i,j,k)}|^2}\,dr\,ds\\
&\quad+C\sum_{i=1}^d\sum_{j=1}^m\int_0^t\int_{\kappa(n,s)}^s\sqrt{\mathbb{E}|\sigma(x_{\kappa(n,r)}^n)-\sigma^n(x_{\kappa(n,r)}^n)|^2\mathbb{E}|\mathbb{T}_r^{(i,j,k)}|^2}\,dr\,ds\\
&\quad+C\int_0^t\sup_{0\le r\le s}\mathbb{E}|e_{r\wedge\nu_{n,R}}^n|^2\,ds+CR^{-\frac{2}{5}}n^{-2},
\end{aligned}
$$

for any $t\in[0,T]$. Then, by (2.19), (2.18), Lemma 2.18, Lemma 2.16, Lemma 2.17, Hölder's inequality and Cauchy-Schwarz inequality, one obtains

$$
\begin{aligned}
T_3(t)+T_4(t) \le\, & Cn^{-1}\mathbb{E}\int_0^t \int_{\kappa(n,s)}^s (1+|x_r|+|x_r^n|)^{2\rho}\,dr \left|\sum_{i=1}^d\sum_{j=1}^m \int_{\kappa(n,s)}^s \mathbb{T}_r^{(i,j,k)}\,dw_r^j\right|^2 ds \\
&+ Cn^{-1}\sum_{i=1}^d\sum_{j=1}^m \int_0^t\int_{\kappa(n,s)}^s (\mathbb{E}(1+|x_r|+|x_r^n|)^{p_0})^{\frac{\rho}{p_0}} \left(\mathbb{E}|\mathbb{T}_r^{(i,j,k)}|^{\frac{2p_0}{p_0-\rho}}\right)^{\frac{p_0-\rho}{p_0}} dr\,ds \\
&+ C\mathbb{E}\int_0^t \sup_{0\le r\le s}\mathbb{E}|e^n_{r\wedge\nu_{n,R}}|^2\,ds + Cn^{-\frac{5+\alpha}{2}} + Cn^{-3} + CR^{-\frac{2}{5}}n^{-2},
\end{aligned}
$$

which, by applying Hölder's inequality, yields

$$
\begin{aligned}
T_3(t)+T_4(t) \le\, & Cn^{-1}\sum_{i=1}^d\sum_{j=1}^m\int_0^t \left(n^{-\frac{p_0}{2\rho}+1}\mathbb{E}\int_{\kappa(n,s)}^s (1+|x_r|+|x_r^n|)^{p_0}\,dr\right)^{\frac{2\rho}{p_0}} \\
&\times \left(n^{-\frac{p_0}{p_0-2\rho}+1}\mathbb{E}\int_{\kappa(n,s)}^s |\mathbb{T}_r^{(i,j,k)}|^{\frac{2p_0}{p_0-2\rho}}\,dr\right)^{\frac{p_0-2\rho}{p_0}} ds \\
&+ C\mathbb{E}\int_0^t \sup_{0\le r\le s}\mathbb{E}|e^n_{r\wedge\nu_{n,R}}|^2\,ds + Cn^{-\frac{5+\alpha}{2}} + CR^{-\frac{2}{5}}n^{-2},
\end{aligned}
$$

for any $t\in[0,T]$. Thus, by using Lemma 2.11 and (2.18), one obtains

$$
T_3(t)+T_4(t) \le C\int_0^t \sup_{0\le r\le s}\mathbb{E}|e^n_{r\wedge\nu_{n,R}}|^2\,ds + Cn^{-\frac{5+\alpha}{2}} + CR^{-\frac{2}{5}}n^{-2},
$$

for any $n\in\mathbb{N}$. Finally, notice that

$$
\begin{aligned}
&\mathbb{E}\int_0^{t\wedge\nu_{n,R}} \left\langle e_s^n, (b(x_s^n)-b(x^n_{\kappa(n,s)})-b_1^n(s,x^n_{\kappa(n,s)})-b_2^n(s,x^n_{\kappa(n,s)}))\right\rangle ds \\
&= \sum_{k=1}^d \mathbb{E}\int_0^{t\wedge\nu_{n,R}} e_s^{n,(k)}(b(x_s^{n,(k)})-b(x^{n,(k)}_{\kappa(n,s)})-b_1^n(s,x^{n,(k)}_{\kappa(n,s)})-b_2^{n,(k)}(s,x^n_{\kappa(n,s)}))\,ds \\
&\le C\int_0^t \sup_{0\le r\le s}\mathbb{E}|e^n_{r\wedge\nu_{n,R}}|^2\,ds + Cn^{-\frac{5+\alpha}{2}} + CR^{-\frac{2}{5}}n^{-2},
\end{aligned}
$$

and the proof is complete. □

Proof of Theorem 2.6. Applying Itô's formula to $|e^n_{t\wedge\nu_{n,R}}|^2$ gives, almost surely,

$$
\begin{aligned}
|e^n_{t\wedge\nu_{n,R}}|^2 =\, & 2\int_0^{t\wedge\nu_{n,R}} \left\langle e_s^n, \bar{b}^n(s,x^n_{\kappa(n,s)})\right\rangle ds + 2\int_0^{t\wedge\nu_{n,R}} \left\langle e_s^n, \bar{\sigma}^n(s,x^n_{\kappa(n,s)})\,dw_s\right\rangle \\
&+ \int_0^{t\wedge\nu_{n,R}} |\bar{\sigma}^n(s,x^n_{\kappa(n,s)})|_{\mathsf{F}}^2\,ds,
\end{aligned}
$$

where $\nu_{n,R}$ is the stopping time defined in (2.17), $\bar{b}^n(t,x^n_{\kappa(n,t)}) = b(x_t)-\tilde{b}^n(t,x^n_{\kappa(n,t)})$ and $\bar{\sigma}^n(t,x^n_{\kappa(n,t)}) = \sigma(x_t)-\tilde{\sigma}^n(t,x^n_{\kappa(n,t)})$. Taking expectations on both sides and using Young's inequality yield, for any $\varepsilon>0$,

$$
\begin{aligned}
\mathbb{E}|e^n_{t\wedge\nu_{n,R}}|^2 \le\, & 2\mathbb{E}\int_0^{t\wedge\nu_{n,R}} \langle e_s^n, (b(x_s)-b(x_s^n))\rangle\,ds \\
&+ 2\mathbb{E}\int_0^{t\wedge\nu_{n,R}} \left\langle e_s^n, (b(x_s^n)-b(x^n_{\kappa(n,s)})-b_1^n(s,x^n_{\kappa(n,s)})-b_2^n(s,x^n_{\kappa(n,s)}))\right\rangle ds
\end{aligned}
$$

$$
\begin{aligned}
&+ 2\mathbb{E}\int_0^{t\wedge\nu_{n,R}} \left\langle e_s^n, (b(x^n_{\kappa(n,s)}) - b^n(x^n_{\kappa(n,s)}))\right\rangle ds \\
&+ (1+\varepsilon)\mathbb{E}\int_0^{t\wedge\nu_{n,R}} |\sigma(x_s) - \sigma(x_s^n)|_{\mathsf{F}}^2\, ds \\
&+ C\mathbb{E}\int_0^{t\wedge\nu_{n,R}} |\sigma(x_s^n) - \sigma(x^n_{\kappa(n,s)}) - \sigma_M^n(s, x^n_{\kappa(n,s)})|_{\mathsf{F}}^2\, ds \\
&+ C\mathbb{E}\int_0^{t\wedge\nu_{n,R}} |\sigma(x^n_{\kappa(n,s)}) - \sigma^n(x^n_{\kappa(n,s)})|_{\mathsf{F}}^2\, ds.
\end{aligned}
$$

for any $t \in [0,T]$. Then, by Cauchy-Schwarz inequality, one obtains

$$
\begin{aligned}
\mathbb{E}|e^n_{t\wedge\nu_{n,R}}|^2 \leq \mathbb{E}&\int_0^{t\wedge\nu_{n,R}} (2e_s^n(b(x_s) - b(x_s^n)) + (1+\varepsilon)|\sigma(x_s) - \sigma(x_s^n)|_{\mathsf{F}}^2)\, ds \\
&+ 2\mathbb{E}\int_0^{t\wedge\nu_{n,R}} \left\langle e_s^n, (b(x_s^n) - b(x^n_{\kappa(n,s)}) - b_1^n(s, x^n_{\kappa(n,s)}) - b_2^n(s, x^n_{\kappa(n,s)}))\right\rangle ds \\
&+ \mathbb{E}\int_0^{t\wedge\nu_{n,R}} |b(x^n_{\kappa(n,s)}) - b^n(x^n_{\kappa(n,s)})|^2\, ds \\
&+ C\mathbb{E}\int_0^{t\wedge\nu_{n,R}} |\sigma(x_s^n) - \sigma(x^n_{\kappa(n,s)}) - \sigma_M^n(s, x^n_{\kappa(n,s)})|^2\, ds \\
&+ C\mathbb{E}\int_0^{t\wedge\nu_{n,R}} |\sigma(x^n_{\kappa(n,s)}) - \sigma^n(x^n_{\kappa(n,s)})|^2\, ds + \int_0^t \sup_{0\leq r\leq s} \mathbb{E}|e^n_{r\wedge\nu_{n,R}}|^2\, ds.
\end{aligned}
$$

Since $p_1 > 2$, applying **A-3** to the first term, and applying Lemma 2.20, 2.16 and 2.17 yield

$$
\sup_{0\leq s\leq t} \mathbb{E}|e^n_{s\wedge\nu_{n,R}}|^2 \leq C\int_0^t \sup_{0\leq r\leq s} \mathbb{E}|e^n_{r\wedge\nu_{n,R}}|^2\, ds + Cn^{-(2+\alpha)} + CR^{-\frac{2}{5}}n^{-2} < \infty,
$$

for any $t \in [0,T]$ and $n \in \mathbb{N}$. Finally, one applies Gronwall's lemma to obtain

$$
\sup_{0\leq s\leq t} \mathbb{E}|e^n_{s\wedge\nu_{n,R}}|^2 \leq Cn^{-(2+\alpha)} + CR^{-\frac{2}{5}}n^{-2},
$$

and the proof is complete by using Fatou's lemma, since the last term in the above inequality vanishes as R tends to infinity.

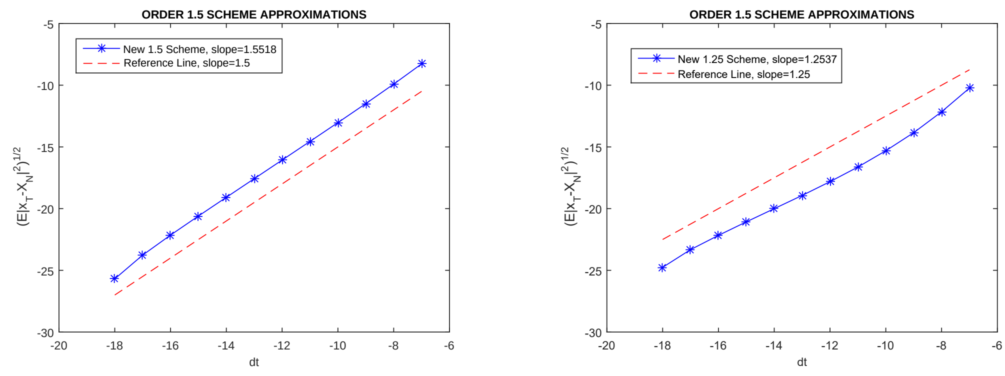
2.5 Simulation results

In this section, simulation results are provided to support the theoretical results in the previous sections. Consider $T = 1$, the step size $\Delta = t_{k+1} - t_k = 1/N$ for $N \in \mathbb{N}$, $t_0 = 0$, and $k \in \{0, \ldots, N-1\}$. For the case $d = m = 1$, the discrete version of the order 1.5 scheme (2.2) is as follows:

$$
\begin{aligned}
X_{k+1} = X_k &+ b^n\Delta + \sigma^n\Delta W + L^{n,1}b\Delta_1 + \frac{1}{2}L^{n,0}b\Delta^2 \\
&+ \frac{1}{2}L^{n,1}\sigma((\Delta W)^2 - \Delta) + L^{n,0}\sigma(\Delta W\Delta - \Delta_1) \\
&+ \frac{1}{2}L^{n,1}L^1\sigma\left(\frac{1}{3}(\Delta W)^2 - \Delta\right)\Delta W,
\end{aligned}
$$

where the following conventions are used:

$$
X_k = X_{t_k}, \quad \Delta W = W_{t_{k+1}} - W_{t_k}, \quad \Delta_1 = \int_{t_k}^{t_{k+1}}\int_{t_k}^{s} dW_r\, ds.
$$



(a) $\alpha = 1$ **(b)** $\alpha = 0.5$

Figure 2.5.1: Rate of convergence of the new order 1.5 scheme with parameters $x_0 = 3$, $c = 0.02$ and $T = 1$. Denote by x_T and X_N respectively the true solution and the numerical approximation of the corresponding SDE at time T. The dashed red lines are the reference lines, and the blue doted lines are the numerical results obtained using the scheme.

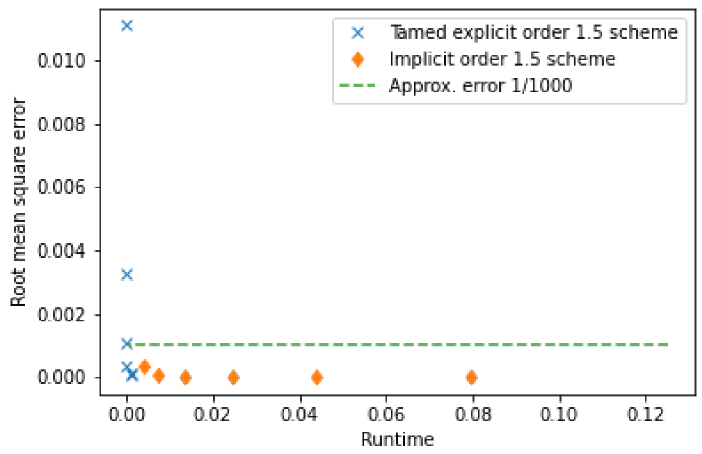


Figure 2.5.2: *Root mean square error (RMSE) of the solutions of the SDE* (2.20) *and of the tamed explicit order 1.5 scheme* (2.2)*, and the RMSE of the solutions of the SDE* (2.20) *and of the implicit order 1.5 scheme as function of runtime when* $N \in \{2^6, 2^7, \dots, 2^{11}\}$.

Note that Δ_1 is normally distributed with mean zero, variance $\frac{1}{3}\Delta^3$, and covariance

$$\mathbb{E}(\Delta_1 \Delta W) = \frac{1}{2}\Delta^2.$$

Then, the following two examples are considered. For the first example, the one-dimensional SDE is given by

$$dx_t = x_t(1 - x_t^2)dt + c(1 - x_t^2)dw_t, \quad \forall t \in [0, T], \tag{2.20}$$

where $T \geq 0$ and $c \in [-0.3086, 0.3086]$. As for the second example, one consider the SDE

$$dx_t = x_t(1 - |x_t|^3)dt + c|x_t|^{\frac{5}{2}}dw_t, \quad \forall t \in [0, T], \tag{2.21}$$

where $T \geq 0$ and $c \in [-0.2209, 0.2209]$. One can check (see Appendix A) that the first example (2.20) satisfies the assumptions **A-1** to **A-5** with $\rho = 2$, whereas the second example (2.21) satisfies the assumptions with $\rho = 4$.

As for the numerical results, Figure 2.5.1 above shows the rate of convergence of the scheme, and the approximations are obtained by simulating 1000 paths. Furthermore, Figure 2.5.1(a)

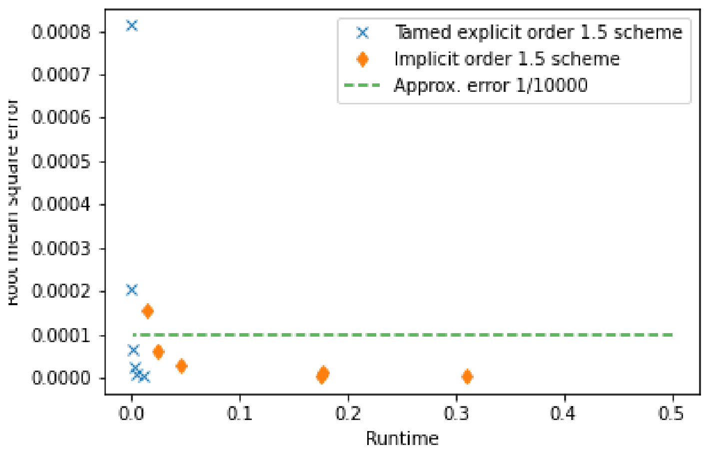


Figure 2.5.3: *Root mean square error (RMSE) of the solutions of the SDE* (2.21) *and of the tamed explicit order 1.5 scheme* (2.2)*, and the RMSE of the solutions of the SDE* (2.21) *and of the implicit order 1.5 scheme as function of runtime when* $N \in \{2^7, 2^8, \ldots, 2^{12}\}$.

illustrates that, for the case $\alpha = 1$, the new explicit order 1.5 scheme has a rate of convergence estimate close to the theoretical result 1.5, which is 1.5518. Similarly, as shown in Figure 2.5.1(b), the slope of the blue line is equal to 1.2537, which supports the theoretical prediction 1.25. Moreover, Figure 2.5.2 shows the root mean square error (RMSE) of the solution of the SDE (2.20) and of the explicit tamed order 1.5 scheme (2.2) and the RMSE of the solution of the SDE (2.20) and of the implicit order 1.5 scheme (see [27, Chapter 12.2]) as a function of runtime when $N \in \{2^5, 2^6, \ldots, 2^{14}\}$. Suppose

$$\left(\sup_{0 \le t \le T} \mathbb{E}|x_t - x_t^n|^2 \right)^{1/2} \le \varepsilon$$

with $\varepsilon = 0.001$. For the first example (2.20), the desired ε precision level is achieved when $N = 2^9$ in the case of the tamed explicit scheme (2.2), while $N = 2^6$ in the case of the implicit scheme. It requires 0.0041 seconds to compute the approximation using the implicit order 1.5 scheme, while it requires 0.0023 seconds using the tamed explicit order 1.5 scheme (2.2) on the same computer. This implies that the tamed explicit order 1.5 scheme (2.2) is 1.78 times faster than the implicit order 1.5 scheme (2.2) on the above computer. Figure 2.5.3 illustrates the RMSE of the solution of the SDE (2.21) and of the explicit tamed order 1.5 scheme (2.2), and the RMSE of the solution of the SDE (2.21) and of the implicit order 1.5 scheme. It takes 0.0241 seconds and 0.0022 seconds for the RMSE to achieve the $\varepsilon = 0.0001$ precision level for the implicit and the tamed explicit scheme, respectively. This implies, in this case, the tamed explicit order 1.5 scheme (2.2) is almost ten times faster than the implicit order 1.5 scheme. The true solutions of the SDE (2.20) and of the SDE (2.21) are obtained by running the implicit (tamed explicit) order 1.5 scheme with $N = 2^{19}$, which are used to compute the approximation errors for the implicit (tamed explicit) order 1.5 scheme. Note that the examples considered in this section are one dimensional. However, in order to implement such algorithms to real world problems where $d \ge 2$, the diffusion coefficient needs to satisfy the commutative condition as discussed in Remark 2.5.

Chapter 3

Higher Order Langevin Monte Carlo Algorithm

3.1 Introduction

In Bayesian statistics and machine learning, one challenge, which has attracted substantial attention in recent years due to its high importance in data-driven applications, is the creation of algorithms which can efficiently sample from a high-dimensional target probability distribution π. We assume π has a density on $\mathbb{R}^d$, denoted by $\hat{\pi}$, such that for any $x \in \mathbb{R}^d$,

$$\hat{\pi}(x) = e^{-U(x)} / \int_{\mathbb{R}^d} e^{-U(y)}\, dy,$$

with $\int_{\mathbb{R}^d} e^{-U(y)}\, dy < \infty$, where $U : \mathbb{R}^d \to \mathbb{R}$ is continuously differentiable. Consider a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$, recall the Langevin SDE associated with π is defined by

$$dx_t = -\nabla U(x_t)dt + \sqrt{2}dw_t, \tag{3.1}$$

where $(w_t)_{t\geq 0}$ is a d-dimensional Brownian motion. In this chapter, we propose a new higher order LMC algorithm (HOLA) to sample from the target distribution π, and it has the following representation, for any $n \in \mathbb{N}$,

$$\overline{X}_{n+1} = \overline{X}_n + \mu_\gamma(\overline{X}_n)\gamma + \sqrt{2\gamma}\sigma_\gamma(\overline{X}_n)\xi_{n+1}, \tag{3.2}$$

where $\gamma \in (0, 1)$ is the step size, $(\xi_n)_{n\in\mathbb{N}}$ are i.i.d. standard d-dimensional Gaussian random variables, for all $x \in \mathbb{R}^d$, $\mu_\gamma(x) = -\nabla U_\gamma(x) + (\gamma/2)(\left(\nabla^2 U \nabla U\right)_\gamma(x) - \vec{\Delta}(\nabla U)_\gamma(x))$, and $\sigma_\gamma(x) = \sqrt{\mathbf{I}_d - \gamma\nabla^2 U_\gamma(x) + (\gamma^2/3)(\nabla^2 U_\gamma(x))^2}$ with $\mathbf{I}_d$ being the $d \times d$ identity matrix. The dependences of the coefficients on γ are given by, for $x \in \mathbb{R}^d$

$$\begin{aligned}
\nabla U_\gamma(x) &= \frac{\nabla U(x)}{(1 + \gamma^{3/2}|\nabla U(x)|^{3/2})^{2/3}}, \\
\nabla^2 U_\gamma(x) &= \frac{\nabla^2 U(x)}{1 + \gamma|\nabla^2 U(x)|}, \\
\left(\nabla^2 U \nabla U\right)_\gamma(x) &= \frac{\nabla^2 U(x)\nabla U(x)}{1 + \gamma|x||\nabla^2 U(x)||\nabla U(x)|}, \\
\vec{\Delta}(\nabla U)_\gamma(x) &= \frac{\vec{\Delta}(\nabla U)(x)}{1 + \gamma^{1/2}|x||\vec{\Delta}(\nabla U)(x)|}.
\end{aligned} \tag{3.3}$$

The tamed coefficients in (3.3) are chosen such that the exponential moments and the desired rate of convergence of the algorithm can be obtained, see Section 3.3 for further discussions. One notes that $\sigma_\gamma^2(x) = \mathbf{I}_d - \gamma\nabla^2 U_\gamma(x) + (\gamma^2/3)(\nabla^2 U_\gamma(x))^2$ is a positive definite matrix. In practice, σ_γ can be computed by generating two independent standard Gaussian noise $\bar{\xi}$ and $\tilde{\xi}$, then one considers $\left(\mathbf{I}_d - (1/2)\gamma\nabla^2 U_\gamma(X_n)\right)\bar{\xi}_{n+1} + (\sqrt{3}/6)\gamma\nabla^2 U_\gamma(x)\tilde{\xi}_{n+1}$, which has the same distribution as $\sigma_\gamma(X_n)\xi_{n+1}$ (see [12] and Chapter 10.4 in [27]). The HOLA algorithm (3.2) is constructed by applying an order 1.5 scheme to the SDE (3.1), see Chapter 10 in [27] for more discussions, which can be expressed explicitly as:

$$\begin{aligned} X_{n+1} = X_n - \nabla U_\gamma(X_n)\gamma + \frac{\gamma^2}{2}\left(\left(\nabla^2 U \nabla U\right)_\gamma(X_n) - \vec{\Delta}(\nabla U)_\gamma(X_n)\right) \\ + \sqrt{2\gamma}\hat{\xi}_{n+1} - \sqrt{2}\nabla^2 U_\gamma(X_n)\dot{\xi}_{n+1} \end{aligned} \tag{3.4}$$

where $(\hat{\xi}_n)_{n\in\mathbb{N}}$ are i.i.d. standard d-dimensional Gaussian random variables, $(\dot{\xi}_n)_{n\in\mathbb{N}}$ are i.i.d. d-dimensional Gaussian random variables with mean $\mathbf{0}$ and covariance $\frac{1}{3}\gamma^3\mathbf{I}_d$ defined by $\dot{\xi}_{n+1} = \int_{t_n}^{t_{n+1}}\int_{t_n}^{s} dw_r\, ds$, and moreover, $\hat{\xi}_n$, $\dot{\xi}_n$ are jointly Gaussian for any $n \in \mathbb{N}$. One notes that for any $n \in \mathbb{N}$, $k, l = 1, \ldots, d$,

$$\mathbb{E}\left(\sqrt{\gamma}\hat{\xi}^{(k)}_{n+1}\dot{\xi}^{(l)}_{n+1}\right) = \begin{cases} \frac{1}{2}\gamma^2, & \text{for } k = l, \\ 0, & \text{otherwise.} \end{cases}$$

One observes that the scheme (3.4) is Markovian, and $\mathcal{L}(X_n)$ is the same as $\mathcal{L}(\overline{X}_n)$, for any $n \in \mathbb{N}$.

This chapter is based on my joint work [46], and it is organised as follows. Section 3.2 presents the assumptions and main results in both super-linear and Lipschitz settings. Section 3.3 discusses the contribution of our work with comparison to the existing literature. In Section 3.4, the proofs of Theorem 3.4 and Theorem 3.5 are provided, while the proofs of Theorem 3.6 and Corollary 3.7 can be found in Section 3.5. An example is provided in Section 3.5.3 illustrating the applicability of the proposed algorithm in the Lipschitz case. Auxiliary results are provided in Appendix B.

3.2 Main results

Assume $U : \mathbb{R}^d \to \mathbb{R}$ is three times continuously differentiable. The following conditions are stated:

B-1 $\liminf_{|x|\to+\infty} |\nabla U(x)| = +\infty$, and $\liminf_{|x|\to+\infty} \frac{\langle x, \nabla U(x)\rangle}{|x||\nabla U(x)|} > 0$.

B-2 There exists $L > 0$, $\rho \geq 2$, and $\alpha \in (0, 1]$, such that for any $i = 1, \ldots, d$ and for all $x, y \in \mathbb{R}^d$,

$$|\nabla^2(\nabla U)^{(i)}(x) - \nabla^2(\nabla U)^{(i)}(y)| \leq L(1 + |x| + |y|)^{\rho-2}|x - y|^\alpha,$$

where $(\nabla U)^{(i)}$ denotes the i-th element of ∇U.

B-3 U is strongly convex, i.e. there exists $m > 0$ such that for all $x, y \in \mathbb{R}^d$,

$$\langle x - y, \nabla U(x) - \nabla U(y)\rangle \geq m|x - y|^2.$$

Remark 3.1. *Unless otherwise specified, the constants $C, K > 0$ may take different values at different places, but these are always independent of the step size $\gamma \in (0, 1)$.*

Remark 3.2. *Assume **B-2** holds, then for any $i = 1, \ldots, d$ and for all $x \in \mathbb{R}^d$,*

$$|\nabla^2(\nabla U)^{(i)}(x)| \leq K(1 + |x|)^{\rho-2+\alpha},$$

where $K = \max\{L, |\nabla^2(\nabla U)^{(i)}(0)|\}$, moreover, for all $x, y \in \mathbb{R}^d$,

$$|\nabla(\nabla U)^{(i)}(x) - \nabla(\nabla U)^{(i)}(y)| \leq K(1 + |x| + |y|)^{\rho-2+\alpha}|x - y|,$$

which implies,

$$|\nabla(\nabla U)^{(i)}(x)| \leq K_1(1+|x|)^{\rho-1+\alpha},$$

where $K_1 = \max\{K, |\nabla(\nabla U)^{(i)}(0)|\}$*. Furthermore, for all* $x, y \in \mathbb{R}^d$*,*

$$|\nabla U^{(i)}(x) - \nabla U^{(i)}(y)| \leq K_1(1+|x|+|y|)^{\rho-1+\alpha}|x-y|,$$

and one obtains

$$|\nabla U^{(i)}(x)| \leq K_2(1+|x|)^{\rho+\alpha},$$

where $K_2 = \max\{K_1, |\nabla U^{(i)}(0)|\}$*. One notes that the above inequality implies*

$$|\vec{\Delta}(\nabla U)(x) - \vec{\Delta}(\nabla U)(y)| \leq d^{3/2}L(1+|x|+|y|)^{\rho-2}|x-y|^{\alpha},$$

$$|\vec{\Delta}(\nabla U)(x)| \leq dK(1+2|x|)^{\rho-1}.$$

Proof. See Appendix B.1 □

Remark 3.3. *By the definition of the tamed coefficients* (3.3) *and* ***B-2****, one obtains for all* $x \in \mathbb{R}^d$*,*

$$|\nabla U_\gamma(x)| \leq \sqrt[3]{2}\gamma^{-1}, \quad |\nabla^2 U_\gamma(x)| \leq \gamma^{-1},$$

$$|\left(\nabla^2 U \nabla U\right)_\gamma(x)| \leq (1+2^{2\rho+1}dK_1K_2)\gamma^{-1},$$

$$|\vec{\Delta}(\nabla U)_\gamma(x)| \leq (1+3^{\rho-1}dK)\gamma^{-1/2}.$$

In particular, when $|x| \geq 1$*,* $x \in \mathbb{R}^d$*, one obtains*

$$|\nabla U_\gamma(x)| \leq \sqrt[3]{2}\gamma^{-1}, \quad |\nabla^2 U_\gamma(x)| \leq \gamma^{-1},$$

$$|\left(\nabla^2 U \nabla U\right)_\gamma(x)| \leq \gamma^{-1}, \quad |\vec{\Delta}(\nabla U)_\gamma(x)| \leq \gamma^{-1/2}.$$

The Markov kernel R_γ associated with (3.2) is given by

$$R_\gamma(x, A) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \mathbb{1}_A\left(x + \mu_\gamma(x)\gamma + \sqrt{2\gamma}\sigma_\gamma(x)y\right) e^{-|y|^2/2}\, dy,$$

for all $x \in \mathbb{R}^d$ and $A \in \mathscr{B}(\mathbb{R}^d)$. Denote by $(P_t)_{t\geq 0}$ the semigroup associated with (3.1). For all $x \in \mathbb{R}^d$ and $A \in \mathscr{B}(\mathbb{R}^d)$, we have $P_t(x, A) = \mathbb{E}[\mathbb{1}_A(x_t)|x_0 = x]$. In addition, for all $x \in \mathbb{R}^d$ and $h \in C^2(\mathbb{R}^d)$, the infinitesimal generator $\mathscr{A}$ associated with (3.1) is defined by

$$\mathscr{A}h(x) = -\langle \nabla U(x), \nabla h(x)\rangle + \Delta h(x).$$

For any $a > 0$, define the Lyapunov function $V_a : \mathbb{R}^d \to [1, +\infty)$ by

$$V_a(x) = \exp\left(a(1+|x|^2)^{1/2}\right), \quad x \in \mathbb{R}^d.$$

Then, for the local Lipschitz drift, one obtains the following convergence results.

Theorem 3.4. *Assume* ***B-1****,* ***B-2*** *and* ***B-3*** *are satisfied. Then, there exist constants* $C > 0$ *and* $\lambda \in (0,1)$ *such that for all* $x \in \mathbb{R}^d$*,* $\gamma \in (0,1)$ *and* $n \in \mathbb{N}$*,*

$$W_2^2(\delta_x R_\gamma^n, \pi) \leq C(\lambda^{n\gamma}V_c(x) + \gamma^{2+\alpha}), \tag{3.5}$$

where c is given in (3.15) *and for all* $\gamma \in (0,1)$*,*

$$W_2^2(\pi_\gamma, \pi) \leq C\gamma^{2+\alpha}.$$

Theorem 3.5. *Assume* ***B-1*** *and* ***B-2*** *are satisfied. There exist* $C > 0$ *and* $\lambda \in (0,1)$ *such that for all* $x \in \mathbb{R}^d$*,* $\gamma \in (0,1)$ *and* $n \in \mathbb{N}$*,*

$$\|\delta_x R_\gamma^n - \pi\|_{V_c^{1/2}} \leq C(\lambda^{n\gamma}V_c(x) + \gamma), \tag{3.6}$$

where c is given in (3.15) *and for all* $\gamma \in (0,1)$,

$$\|\pi_\gamma - \pi\|_{V_c^{1/2}} \leq C\gamma. \tag{3.7}$$

In the case of super-linear coefficients, tracking the explicit constants involves tedious calculations, and it is less informative compared to the case of Lipschitz coefficients, in the sense that the dependence on the dimension of the constant C (appearing in Theorem 3.4 and Theorem 3.5) is $O(e^d)$. One notes that this is due to the fact that exponential moments of the scheme 3.2 is obtained when a log-Sobolev inequality is used. To illustrate the explicit dependence on the dimension, and to provide explicit constants for the moment bounds and the convergence in Wasserstein distance, the Lipschitz case is considered. Four times continuous differentiablility on U is required and the following conditions are assumed:

B-4 There exists $L_1 > 0$, such that for all $x, y \in \mathbb{R}^d$,

$$|\nabla U(x) - \nabla U(y)| \leq L_1|x-y|.$$

B-5 There exists $L_2 > 0$, such that for all $x, y \in \mathbb{R}^d$,

$$|\nabla^2 U(x) - \nabla^2 U(y)| \leq L_2|x-y|.$$

B-6 There exists $L > 0$, such that for all $x, y \in \mathbb{R}^d$,

$$|\nabla^2 (\nabla U)^{(i)}(x) - \nabla^2 (\nabla U)^{(i)}(y)| \leq L|x-y|.$$

One notices that, in the Lipschitz case, there is no need to use the tamed coefficients, and one can consult Theorem 10.6.3 in [27] for the classical strong convergence result for the order 1.5 scheme in a finite time. The counterpart of algorithm (3.2) in the Lipschitz case becomes: for any $n \in \mathbb{N}$

$$\tilde{X}_{n+1} = \tilde{X}_n + \mu(\tilde{X}_n)\gamma + \sqrt{2\gamma}\sigma(\tilde{X}_n)\xi_{n+1}, \tag{3.8}$$

where $\gamma \in (0,1)$ is the step size, $(\xi_n)_{n\in\mathbb{N}}$ are i.i.d. standard d-dimensional Gaussian random variables, for all $x \in \mathbb{R}^d$, $\mu(x) = -\nabla U(x) + (\gamma/2)(\nabla^2 U(x)\nabla U(x) - \vec{\Delta}(\nabla U)(x))$, and $\sigma(x) = \sqrt{\mathbf{I}_d - \gamma\nabla^2 U(x) + (\gamma^2/3)(\nabla^2 U(x))^2}$. The Markov kernel $\tilde{R}_\gamma$ associated with (3.8) is given by

$$\tilde{R}_\gamma(x, A) = (2\pi)^{-d/2}\int_{\mathbb{R}^d} \mathbb{1}_A\left(x + \mu(x)\gamma + \sqrt{2\gamma}\sigma(x)y\right)e^{-|y|^2/2}\,dy,$$

for all $x \in \mathbb{R}^d$ and $A \in \mathscr{B}(\mathbb{R}^d)$

Theorem 3.6. *Let* $\gamma \in \left(0, \frac{1}{\tilde{m}} \wedge \frac{4}{5(m+L_1)} \wedge \frac{\sqrt{m}}{(m+L_1)\sqrt{L_1}} \wedge \sqrt[3]{\frac{1}{(m+L_1)L_1^2}}\right)$. *Assume **B-3** - **B-6** are satisfied. Then, for all* $x \in \mathbb{R}^d$ *and* $n \in \mathbb{N}$,

$$W_2^2(\delta_x \tilde{R}_\gamma^n, \pi) \leq e^{-mn\gamma}\left(2|x-x^*|^2 + \frac{2d}{m}\right) + \bar{C}\gamma^3,$$

where $\tilde{m}$ *is given in* (3.35), $\bar{C} = O(d^4)$ *and its the explicit expression is given in the proof.*

Corollary 3.7. *Let* $\gamma \in \left(0, \frac{1}{\tilde{m}} \wedge \frac{4}{5(m+L_1)} \wedge \frac{\sqrt{m}}{(m+L_1)\sqrt{L_1}} \wedge \sqrt[3]{\frac{1}{(m+L_1)L_1^2}}\right)$. *Assume **B-3** - **B-6** are satisfied. If one considers a multivariate Gaussian as the target distribution, then for all* $x \in \mathbb{R}^d$ *and* $n \in \mathbb{N}$,

$$W_2^2(\delta_x \tilde{R}_\gamma^n, \pi) \leq e^{-mn\gamma}\left(2|x-x^*|^2 + \frac{2d}{m}\right) + \tilde{C}\gamma^3.$$

where $\tilde{m}$ *is given in* (3.35), $\tilde{C} = O(d)$ *and its the explicit expression is given in the proof.*

Remark 3.8. *One notices that only three times continuous differentiablility on the potential* U *is required in the case of super-linear coefficients, while we assume four times continuous differentiablility in the case of Lipschitz coefficients. This further smoothness in the Lipschitz* *case is reuqired in order to obtain a better dependence on the dimension of the bound in Wasserstein distance, i.e. to obtain $\bar{C} = O(d^4)$ in Theorem 3.6. While one can still obtain similar results in Theorem 3.6 and Corollary 3.7 without assuming further smoothness, the dependence on dimension of the bound will increase to $O(d^6)$.*

3.3 Related work and discussion

Higher order scheme. The higher order LMC algorithm (3.2) is obtained using the Itô-Taylor (Wagner-Platen) expansion, see [41] and Section 10.4 in [27]. It is suggested in Section 10.6 in [27] that any higher order schemes can be constructed using such an approach. One notices that the LMCO' algorithm considered in [12], which is obtained using the LMC algorithm with the Ozaki discretization, is close to the algorithm (3.8), which is the counterpart of the algorithm (3.2) in the Lipschitz case. The difference between the two algorithms is that there is one more term $\vec{\Delta}(\nabla U)$ in (3.8). Without this term, the rate of convergence of the algorithm (3.8) reduces from 1.5 to 1 in the Wasserstein-2 distance.
Tamed coefficients. The algorithm (3.2) of the SDE (3.1) with superlinear coefficient is constructed using a taming technique, which is first introduced in [25] for the Euler scheme and is further developed in [44]. Then, a uniform taming approach is suggested in [30] which allows natural extensions of the taming technique to higher order schemes. In other words, it suggests that each coefficient in the numerical scheme should be multiplied by the same taming factor (see Remark 2 in [30]). However, in this chapter, different terms in the algorithm (3.2) have different taming factors as defined in (3.3). The reason is that, instead of a direct application of Itô's formula, one uses the derivation of the log-sobolev inequality to obtain exponential moment bounds for the numerical scheme (3.2) in an infinite time horizon (see Proposition 3.10 for a detailed proof). This requires an additional assumption **B-1**. Moreover, the choice of the taming factors is crucial in the sense that the tamed coefficients should converge to the original coefficients with a desired rate.
Rate of convergence. In the context of SDEs with superlinear coefficients, the strong convergence results of the tamed numerical schemes have been studied in depth in literature. One may refer to [5], [25], [30], [44], [45], [50] for the convergence resutls of tamed Euler and Milstein schemes in a finite time. In addition, Theorem 1 in [47] provides a strong convergence result in $\mathscr{L}^2$ of the tamed order 1.5 scheme. While the aforementioned results focused on the convergence rates in finite time horizons, [7] considers a TULA algorithm which provides rate 1 in Wasserstein-2 distance and rate 1/2 in total variation. By extending the results in [47] and [7], Theorem 3.4 and Theorem 3.5 state that the convergence results of the HOLA algorithm (3.2) in Wasserstein-2 distance and in total variation can be improved to rate $1 + \alpha/2$ and rate 1 respectively. One notices that the assumptions **B-1** and **B-3** are the same as the assumptions in [7], while the local Hölder condition **B-2** is the same as the assumption A-4 in [47].
As for the SDEs with Lipschitz coefficients, [11], [12], [15] and [14] provide convergence results in Wasserstein-2 distance and in total variation for the ULA algorithm. In addition, LMCO and LMCO' algorithms are considered in [11] and [12] which make use of the Hessian of U, however, the rate of convergence is shown to have the same order as ULA in Wasserstein-2 distance. Under **B-3** - **B-6**, Theorem 3.6 provides a convergence result for the scheme (3.2) in Wasserstein-2 distance, which is of order 1.5. It improves existing results by imposing four times differentiability on the potential U and an additional assumption **B-6**.
Non-asymptotic bounds and computational complexity. The nonasymptotic bounds in total variation between the ULA algorithm and SDE (3.1) are established in [11]. Subsequently, improved results, including the Wasserstein-2 distance, are provided in [12], [15] and [14] with better dependence on the dimension. Theorem 3.6 in this chapter provides the non-asymptotic bound between the HOLA algorithm (3.2) and the target distribution π in Wasserstein-2 distance for the Lipschitz case. It shows that the dependence on dimension is $O(d^4)$, and the number of iterations required to reach ε percision level is given precisely by

$$n \geq \left((2\bar{C})^{\frac{1}{3}}/m\varepsilon^{\frac{2}{3}}\right) \log\left(4(|x - x^*|^2 + d/m)/\varepsilon^2\right)$$

with $\bar{C} = O(d^4)$. This implies that compared to results in [12] and [15], the HOLA algorithm

(3.2) requires fewer steps to reach a suitably high precision level, i.e. for $\varepsilon < O(d^{-1})$. As for the computational complexity of the algorithm (3.2), it shows in [19] that the computational cost for the Hessian-vector product is not more expensive than evaluating the gradient. Moreover, although the computational cost for one iteration increases due to third derivatives of U, there are techniques which can be employed to reduce the computational cost dramatically, see [20], [21] and references therein.

3.4 Local Lipschitz case

3.4.1 Moment bounds

It is a well-known result that by **B-1** and **B-2**, the SDE (3.1) has a unique strong solution. One then needs to obtain moment bounds of the SDE (3.1) and the numerical scheme (3.2) before considering the convergence results.

By using Foster-Lyapunov conditions, one can obtain the exponential moment bounds for the solution of SDE (3.1). More concretely, the application of Theorem 1.1, 6.1 in [43] and Theorem 2.2 in [38] yields the following results.

Proposition 3.9. *Assume **B-1** and **B-2** are satisfied. For all $a > 0$, there exists $b_a > 0$, such that for all $x \in \mathbb{R}^d$,*

$$\mathscr{A} V_a(x) \leq -aV_a(x) + ab_a,$$

and

$$\sup_{t\geq 0} P_t V_a(x) \leq V_a(x) + b_a.$$

Furthermore, there exist $C_a > 0$ and $\rho_a \in (0,1)$ such that for all $t > 0$ and probability measures μ_0, ν_0 on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ satisfying $\mu_0(V_a) + \nu_0(V_a) < +\infty$,

$$\|\mu_0 P_t - \nu_0 P_t\|_{V_a} \leq C_a \rho_a^t \|\mu_0 - \nu_0\|_{V_a}, \quad \|\mu_0 P_t - \pi\|_{V_a} \leq C_a \rho_a^t \mu_0(V_a).$$

Proof. One can refer to Proposition 1 in [7] for the detailed proof. □

The proposition below provides a uniform bound for exponential moments of the Markov chain $(\overline{X}_k)_{k\geq 0}$.

Proposition 3.10. *Assume **B-1** and **B-2** are satisfied. Then, there exist constants $b, c, M > 0$, such that for all $x \in \mathbb{R}^d$ and $\gamma \in (0,1)$,*

$$R_\gamma V_c(x) \leq e^{-\frac{7}{3}c^2\gamma} V_c(x) + \gamma b \mathbb{1}_{\overline{\mathrm{B}}(0,M)}(x),$$

and for all $n \in \mathbb{N}$

$$R_\gamma^n V_c(x) \leq e^{-\frac{7}{3}c^2 n\gamma} V_c(x) + \frac{3b}{7c^2} e^{\frac{7}{3}c^2\gamma}.$$

Moreover, this guarantees that the Gaussian kernel R_γ has a unique invariant measure π_γ and R_γ is geometrically ergodic w.r.t. π_γ.

Proof. We use the scheme (3.2) throughout the proof. First, one observes that by **B-1**, for $\gamma \in (0,1)$, the following holds

$$\liminf_{|x|\to+\infty} \left\langle \frac{x}{|x|}, \nabla U_\gamma(x) \right\rangle - \frac{\gamma}{2|x|}|\nabla U_\gamma(x)|^2 > 0. \tag{3.9}$$

Indeed, by **B-1**, there exist $M', \kappa > 0$ such that for all $|x| \geq M'$, $x \in \mathbb{R}^d$, $\langle x, \nabla U(x)\rangle \geq \kappa |x||\nabla U(x)|$. Then, we have for all $|x| \geq M'$, $x \in \mathbb{R}^d$,

$$\begin{aligned}
&\left\langle \frac{x}{|x|}, \nabla U_\gamma(x) \right\rangle - \frac{\gamma}{2|x|}|\nabla U_\gamma(x)|^2 \\
&\geq \frac{1}{2|x|(1+\gamma^{3/2}|\nabla U(x)|^{3/2})^{2/3}} \left(2\kappa|x||\nabla U(x)| - \frac{\gamma|\nabla U(x)|^2}{(1+\gamma^{3/2}|\nabla U(x)|^{3/2})^{2/3}} \right)
\end{aligned}$$

$$
\begin{aligned}
&\geq \frac{|\nabla U(x)|}{2|x|(1+\gamma^{3/2}|\nabla U(x)|^{3/2})^{2/3}} \left(2\kappa|x| - \frac{\sqrt[3]{2}\gamma|\nabla U(x)|}{1+\gamma|\nabla U(x)|}\right) \\
&\geq \frac{|\nabla U(x)|}{2(1+\gamma^{3/2}|\nabla U(x)|^{3/2})^{2/3}} \left(2\kappa - \frac{\sqrt[3]{2}}{|x|}\right).
\end{aligned}
$$

Meanwhile, by **B-1**, there exist $M'', K > 0$ such that for all $|x| \geq M''$, $x \in \mathbb{R}^d$, $|\nabla U| \geq K$. Note that $f(x) = x/(1+x^{3/2})^{2/3}$ is a non-decreasing function for all $x \geq 0$. Then, one obtains (3.9), since for all $x \in \mathbb{R}^d$, $|x| \geq \max(M', M'', \sqrt[3]{2}\kappa^{-1})$

$$
\left\langle \frac{x}{|x|}, \nabla U_\gamma(x) \right\rangle - \frac{\gamma}{2|x|}|\nabla U_\gamma(x)|^2 \geq \frac{\kappa K}{2(1+\gamma^{3/2}K^{3/2})^{2/3}} > 0.
$$

The function $f(x) = (1+|x|^2)^{1/2}$ is Lipschitz continuous with Lipschitz constant equal to 1. Let $\overline{X}_0 = x$, then for all $x \in \mathbb{R}^d$, applying log Sobolev inequality (see Proposition 5.5.1 in [1] and Appendix B.2 for a detailed proof) gives,

$$
R_\gamma V_a(x) = \mathbb{E}_x(V_a(\overline{X}_1)) \leq e^{\frac{7}{3}\gamma a^2} \exp\left\{a\mathbb{E}((1+|\overline{X}_1|^2)^{1/2}|\overline{X}_0 = x)\right\}, \tag{3.10}
$$

which using Jensen's inequality yields

$$
R_\gamma V_a(x) \leq e^{\frac{7}{3}\gamma a^2} \exp\left\{a\left(1+\mathbb{E}\left(\left|\overline{X}_0 + \mu_\gamma(\overline{X}_0)\gamma + \sigma_\gamma(\overline{X}_0)\sqrt{2\gamma}\xi_1\right|^2 \middle| \overline{X}_0 = x\right)\right)^{1/2}\right\}. \tag{3.11}
$$

One calculates

$$
\begin{aligned}
\mathbb{E}\left[\left|\sigma_\gamma(\overline{X}_0)\sqrt{2\gamma}\xi_1\right|^2 \middle| \overline{X}_0 = x\right] &\leq 2\gamma\left|\sigma_\gamma(x)\right|^2 \mathbb{E}\left[|\xi_1|^2\right] \\
&\leq 2\gamma d + \frac{2\gamma^3}{3}\left|\nabla^2 U_\gamma(x)\right|^2 d + 2\gamma^2\left|\nabla^2 U_\gamma(x)\right| d \\
&\leq \frac{14}{3}d\gamma.
\end{aligned} \tag{3.12}
$$

Then, by inserting (3.12) into (3.11), one obtains

$$
R_\gamma V_a(x) \leq e^{\frac{7}{3}\gamma a^2} \exp\left\{a\left(1 + A_\gamma(x) + \frac{14}{3}d\gamma\right)^{1/2}\right\}, \tag{3.13}
$$

where

$$
A_\gamma(x) = \left|x - \nabla U_\gamma(x)\gamma + \frac{\gamma^2}{2}\left(\left(\nabla^2 U \nabla U\right)_\gamma(x) - \vec{\Delta}(\nabla U)_\gamma(x)\right)\right|^2.
$$

Then, expanding the square yields

$$
\begin{aligned}
A_\gamma(x) = &|x|^2 - 2\gamma\langle x, \nabla U_\gamma(x)\rangle + \gamma^2\left|\nabla U_\gamma(x)\right|^2 - \gamma^2\left\langle x, \vec{\Delta}(\nabla U)_\gamma(x)\right\rangle \\
&+ \frac{\gamma^4}{4}\left|\vec{\Delta}(\nabla U)_\gamma(x)\right|^2 + \gamma^2\left\langle x, \left(\nabla^2 U \nabla U\right)_\gamma(x)\right\rangle \\
&- \gamma^3\left\langle \nabla U_\gamma(x), \left(\nabla^2 U \nabla U\right)_\gamma(x)\right\rangle + \gamma^3\left\langle \nabla U_\gamma(x), \vec{\Delta}(\nabla U)_\gamma(x)\right\rangle \\
&+ \frac{\gamma^4}{4}\left|\left(\nabla^2 U \nabla U\right)_\gamma(x)\right|^2 - \frac{\gamma^4}{2}\left\langle \left(\nabla^2 U \nabla U\right)_\gamma(x), \vec{\Delta}(\nabla U)_\gamma(x)\right\rangle.
\end{aligned}
$$

By (3.9), there exist $M_1, \kappa_1 > 0$ such that for all $|x| \geq M_1$,

$$
\langle x, \nabla U_\gamma(x)\rangle - \frac{\gamma}{2}|\nabla U_\gamma(x)|^2 > \kappa_1|x|.
$$

Thus, by using Remark 3.3, for all $|x| \geq \max\{1, M_1\}$,

$$\begin{aligned}A_\gamma(x) + \frac{14}{3}d\gamma &\leq |x|^2 - 2\gamma\kappa_1|x| + \gamma^{3/2} + \frac{1}{4}\gamma^3 \\ &\quad + 3\gamma + 2\gamma^{3/2} + \frac{1}{4}\gamma^2 + \frac{1}{2}\gamma^{5/2} + \frac{14}{3}d\gamma \\ &\leq |x|^2 - 2\gamma\kappa_1|x| + \frac{35}{3}d\gamma.\end{aligned}$$

Denote by $M = \max\{1, M_1, \frac{35}{3}d(\kappa_1)^{-1}\}$, for all $x \in \mathbb{R}^d$, $|x| \geq M$,

$$A_\gamma(x) + \frac{14}{3}d\gamma \leq |x|^2 - \gamma\kappa_1|x|.$$

For $t \in [0,1]$, $(1-t)^{1/2} \leq 1 - t/2$ and $g(x) = x/(1+x^2)^{1/2}$ is a non-decreasing function for all $x \geq 0$. Then, for all $x \in \mathbb{R}^d$, $|x| \geq M$

$$\begin{aligned}\left(1 + A_\gamma(x) + \frac{14}{3}d\gamma\right)^{1/2} &\leq \left(1+|x|^2\right)^{1/2}\left(1 - \frac{7\gamma}{3}\frac{3\kappa_1|x|}{7(1+|x|^2)}\right)^{1/2} \\ &\leq \left(1+|x|^2\right)^{1/2} - \frac{7\gamma}{3}\frac{3\kappa_1 M}{14(1+M^2)^{1/2}}.\end{aligned} \tag{3.14}$$

By substituting (3.14) into (3.13) and completing the square, one obtains, for $|x| \geq M$,

$$R_\gamma V_c(x) \leq e^{-\frac{7}{3}c^2\gamma}V_c(x),$$

where

$$c = \frac{3\kappa_1 M}{28(1+M^2)^{1/2}}. \tag{3.15}$$

For the case $|x| \leq M$, by Remark 3.3, the following result can be obtained:

$$A_\gamma(x) \leq |x|^2 + c^\sharp\gamma(1+M)^{4\rho+2},$$

where $c^\sharp$ is a positive constant (that depends on d and L). Then, by using $(1+s_1+s_2)^{1/2} \leq (1+s_1)^{1/2} + s_2/2$ for $s_1, s_2 \geq 0$,

$$\left(1 + A_\gamma(x) + \frac{14}{3}d\gamma\right)^{1/2} \leq (1+|x|^2)^{1/2} + \gamma\left(\frac{c^\sharp}{2}(1+M)^{4\rho+2} + \frac{7d}{3}\right).$$

Thus,

$$R_\gamma V_c(x) \leq e^{\theta\gamma}V_c(x),$$

where $\theta = \frac{7}{3}c^2 + c(\frac{c^\sharp}{2}(1+M)^{4\rho+2} + \frac{7}{3}d)$. Moreover, for $|x| \leq M$,

$$R_\gamma V_c(x) - e^{-\frac{7}{3}c^2\gamma}V_c(x) \leq e^{\theta\gamma}(1 - e^{-\gamma(\frac{7}{3}c^2+\theta)})V_c(x) \leq \gamma e^{\theta\gamma}\left(\frac{7}{3}c^2 + \theta\right)V_c(x).$$

Denote by $b = e^{(\theta\gamma + c\sqrt{1+M^2})}\left(\frac{7}{3}c^2 + \theta\right)$, one obtains

$$R_\gamma V_c(x) \leq e^{-\frac{7}{3}c^2\gamma}V_c(x) + \gamma b \mathbb{1}_{\overline{\mathrm{B}}(0,M)}(x).$$

Then by induction, for all $n \in \mathbb{N}$ and $x \in \mathbb{R}$

$$\begin{aligned}R_\gamma^n V_c(x) &\leq e^{-\frac{7}{3}c^2 n\gamma}V_c(x) + \frac{1 - e^{-\frac{7}{3}c^2 n\gamma}}{1 - e^{-\frac{7}{3}c^2\gamma}}\gamma b \\ &\leq e^{-\frac{7}{3}c^2 n\gamma}V_c(x) + \frac{3b}{7dc^2}e^{\frac{7}{3}c^2\gamma},\end{aligned}$$

the last inequality holds since $1-e^{-\frac{7}{3}c^2\gamma}=\int_0^\gamma \frac{7}{3}c^2 e^{-\frac{7}{3}c^2 s}\,ds \geq \frac{7}{3}c^2\gamma e^{-\frac{7}{3}c^2\gamma}$. Finally, since any compact set on $\mathbb{R}^d$ is accessible and small for R_γ, then by section 3.1 in [43] and Theorem 15.0.1 in [39], for all $\gamma \in (0,1)$, R_γ has a unique invariant measure π_γ and it is geometrically ergodic w.r.t. π_γ. $\square$

The results in Proposition 3.9 and 3.10 provide exponential moment bounds for the solution of SDE (3.1) and the scheme (3.2), which enable us to consider the total variation and Wasserstein distance between the target distribution π and the n-th iteration of the MCMC algorithm.

3.4.2 Proof of Theorem 3.4

In order to obtain the convergence rate in Wasserstein distance, the assumption **B-3** is needed, which assumes the convexity of U. We consider the linear interpolation of the scheme (3.4) given by

$$\bar{x}_t = \bar{x}_0 - \int_0^t \nabla\tilde{U}_\gamma(s, \bar{x}_{\lfloor s/\gamma\rfloor\gamma})\,ds + \sqrt{2}w_t, \tag{3.16}$$

for all $t \geq 0$, where

$$\nabla\tilde{U}_\gamma(s,\bar{x}_{\lfloor s/\gamma\rfloor\gamma}) = \nabla U_\gamma(\bar{x}_{\lfloor s/\gamma\rfloor\gamma}) + \nabla U_{1,\gamma}(s,\bar{x}_{\lfloor s/\gamma\rfloor\gamma}) + \nabla U_{2,\gamma}(s,\bar{x}_{\lfloor s/\gamma\rfloor\gamma}),$$

with

$$\nabla U_{1,\gamma}(s,\bar{x}_{\lfloor s/\gamma\rfloor\gamma}) = -\int_{\lfloor s/\gamma\rfloor\gamma}^s \left((\nabla^2 U\nabla U)_\gamma(\bar{x}_{\lfloor s/\gamma\rfloor\gamma}) - \vec{\Delta}(\nabla U)_\gamma(\bar{x}_{\lfloor s/\gamma\rfloor\gamma})\right) dr,$$

$$\nabla U_{2,\gamma}(s,\bar{x}_{\lfloor s/\gamma\rfloor\gamma}) = \sqrt{2}\int_{\lfloor s/\gamma\rfloor\gamma}^s \nabla^2 U_\gamma(\bar{x}_{\lfloor s/\gamma\rfloor\gamma})\,dw_r.$$

Note that the linear interpolation (3.16) and the scheme (3.4) coincide at grid points, i.e. for any $n \in \mathbb{N}$, $X_n = \bar{x}_{n\gamma}$. Let $(\mathscr{F}_t)_{t\geq 0}$ be a filtration associated with $(w_t)_{t\geq 0}$. For any $n \in \mathbb{N}$, denote by $\mathbb{E}^{\mathscr{F}_{n\gamma}}[\cdot]$ the expectation conditional on $\mathscr{F}_{n\gamma}$.

Lemma 3.11. *Assume* ***B-1*** *and* ***B-2*** *are satisfied. Then, there exists a constant $C > 0$ such that for all $p > 0$, $\gamma \in (0,1)$, $n \in \mathbb{N}$, and $t \in [n\gamma,(n+1)\gamma)$,*

$$\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\nabla U_{1,\gamma}(t,\bar{x}_{n\gamma})|^p\right] \leq C\gamma^p V_c(\bar{x}_{n\gamma}),$$

$$\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\nabla U_{2,\gamma}(t,\bar{x}_{n\gamma})|^p\right] \leq C\gamma^{\frac{p}{2}} V_c(\bar{x}_{n\gamma}).$$

Proof. Consider a polynomial function $f(|x|) \in C_{poly}(\mathbb{R}_+,\mathbb{R}_+)$, then there exists a constant $C > 0$ such that for all $x \in \mathbb{R}^d$, $f(|x|) \leq CV_c(x)$. For $p > 1$, by applying Hölder's inequality and Remark 3.2, one obtains

$$\begin{aligned}
&\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\nabla U_{1,\gamma}(t,\bar{x}_{n\gamma})|^p\right] \\
&= \mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left|-\int_{n\gamma}^t \left((\nabla^2 U\nabla U)_\gamma(\bar{x}_{n\gamma}) - \vec{\Delta}(\nabla U)_\gamma(\bar{x}_{n\gamma})\right) dr\right|^p\right] \\
&\leq C\gamma^{p-1}\int_{n\gamma}^t \mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left|(\nabla^2 U\nabla U)_\gamma(\bar{x}_{n\gamma})\right|^p\right] dr \\
&\quad + C\gamma^{p-1}\int_{n\gamma}^t \mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left|\vec{\Delta}(\nabla U)_\gamma(\bar{x}_{n\gamma})\right|^p\right] dr \\
&\leq C\gamma^p V(\bar{x}_{n\gamma}),
\end{aligned}$$

The second inequality can be proved using similar arguments. For the case $0 < p \leq 1$, Jensen's inequality is used to obtain the desired result. $\square$

Lemma 3.12. *Assume **B-1** and **B-2** are satisfied. Then, there exists a constant $C > 0$ such that for all $p > 0$, $\gamma \in (0,1)$, $n \in \mathbb{N}$, and $t \in [n\gamma, (n+1)\gamma)$,*

$$\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\bar{x}_t - \bar{x}_{n\gamma}|^p\right] \leq C\gamma^{\frac{p}{2}} V_c(\bar{x}_{n\gamma}),$$

$$\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|x_t - x_{n\gamma}|^p\right] \leq C\gamma^{\frac{p}{2}} V_c(x_{n\gamma}).$$

Proof. For $p > 1$, by using Hölder's inequality, Remark 3.2 and Lemma 3.11, we have

$$\begin{aligned}
&\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\bar{x}_t - \bar{x}_{n\gamma}|^p\right] \\
&= \mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left|-\int_{n\gamma}^t \nabla\tilde{U}_\gamma(s, \bar{x}_{n\gamma})\,ds + \sqrt{2}\int_{n\gamma}^t dw_s\right|^p\right] \\
&\leq C\gamma^{p-1}\int_{n\gamma}^t \mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\nabla U_\gamma(\bar{x}_{n\gamma}) + \nabla U_{1,\gamma}(s, \bar{x}_{n\gamma}) + \nabla U_{2,\gamma}(s, \bar{x}_{n\gamma})|^p\right]ds + C\gamma^{\frac{p}{2}} \\
&\leq C\gamma^{\frac{p}{2}} V_c(\bar{x}_{n\gamma}).
\end{aligned}$$

For the case $0 < p \leq 1$, one can use Jensen's inequality to obtain

$$\begin{aligned}
\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\bar{x}_t - \bar{x}_{n\gamma}|^p\right] &\leq \left(\mathbb{E}^{\mathscr{F}_{n\gamma}}\left|\int_{n\gamma}^t \nabla\tilde{U}_\gamma(s, \bar{x}_{n\gamma})\,ds + \sqrt{2}\int_{n\gamma}^t dw_s\right|\right)^p \\
&\leq C\gamma^{\frac{p}{2}} V_c(\bar{x}_{n\gamma}),
\end{aligned}$$

Similarly, for $p > 1$, by using Hölder's inequality, one obtains

$$\begin{aligned}
\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|x_t - x_{n\gamma}|^p\right] &= \mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left|-\int_{n\gamma}^t \nabla U(x_s)\,ds + \sqrt{2}\int_{n\gamma}^t dw_s\right|^p\right] \\
&\leq C\gamma^{p-1}\int_{n\gamma}^t \mathbb{E}^{\mathscr{F}_{n\gamma}}\left(1 + |x_s|^{p(\rho+\alpha)}\right)ds + C\gamma^{\frac{p}{2}} \\
&\leq C\gamma^{\frac{p}{2}} V_c(x_{n\gamma}),
\end{aligned}$$

where the last inequality holds due to Proposition 3.9. The case $p \in (0,1]$ follows from the application of Jensen's inequality. ☐

Lemma 3.13. *Assume **B-1** and **B-2** are satisfied. Then, there exists a constant $C > 0$ such that for all $\gamma \in (0,1)$, $n \in \mathbb{N}$, and $t \in [n\gamma, (n+1)\gamma)$,*

$$\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\nabla U(\bar{x}_t) - \nabla U(\bar{x}_{n\gamma}) - \nabla U_{1,\gamma}(t, \bar{x}_{n\gamma}) - \nabla U_{2,\gamma}(t, \bar{x}_{n\gamma})|^2\right] \leq C\gamma^2 V_c(\bar{x}_{n\gamma}).$$

Proof. For any $t \in [n\gamma, (n+1)\gamma)$, applying Itô's formula to $\nabla U(\bar{x}_t)$ gives, almost surely

$$\begin{aligned}
&\nabla U(\bar{x}_t) - \nabla U(\bar{x}_{n\gamma}) \\
&= -\int_{n\gamma}^t \left(\nabla^2 U(\bar{x}_r)\nabla\tilde{U}_\gamma(r, \bar{x}_{n\gamma}) - \vec{\Delta}(\nabla U)(\bar{x}_r)\right)dr + \sqrt{2}\int_{n\gamma}^t \nabla^2 U(\bar{x}_r)\,dw_r \\
&= -\int_{n\gamma}^t \left(\nabla^2 U(\bar{x}_r) - \nabla^2 U(\bar{x}_{n\gamma})\right)\nabla U_\gamma(\bar{x}_{n\gamma})\,dr - \int_{n\gamma}^t \nabla^2 U(\bar{x}_{n\gamma})\nabla U_\gamma(\bar{x}_{n\gamma})\,dr \\
&\quad - \int_{n\gamma}^t \nabla^2 U(\bar{x}_r)(\nabla U_{1,\gamma}(r, \bar{x}_{n\gamma}) + \nabla U_{2,\gamma}(r, \bar{x}_{n\gamma}))\,dr \\
&\quad + \sqrt{2}\int_{n\gamma}^t \left(\nabla^2 U(\bar{x}_r) - \nabla^2 U(\bar{x}_{n\gamma})\right)dw_r + \sqrt{2}\int_{n\gamma}^t \nabla^2 U(\bar{x}_{n\gamma})\,dw_r \\
&\quad + \int_{n\gamma}^t \left(\vec{\Delta}(\nabla U)(\bar{x}_r) - \vec{\Delta}(\nabla U)(\bar{x}_{n\gamma})\right)dr + \int_{n\gamma}^t \vec{\Delta}(\nabla U)(\bar{x}_{n\gamma})\,dr.
\end{aligned}$$

By substracting $\nabla U_{1,\gamma}(t, \bar{x}_{n\gamma})$, $\nabla U_{2,\gamma}(t, \bar{x}_{n\gamma})$, squaring both sides and taking conditional expectation yields,

$$\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\nabla U(\bar{x}_t)-\nabla U(\bar{x}_{n\gamma})-\nabla U_{1,\gamma}(t,\bar{x}_{n\gamma})-\nabla U_{2,\gamma}(t,\bar{x}_{n\gamma})|^2\right]\leq C\sum_{i=1}^{5}G_i(t). \tag{3.17}$$

where

$$\begin{aligned}
G_1(t)&=\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left|-\int_{n\gamma}^{t}\left(\nabla^2U(\bar{x}_r)-\nabla^2U(\bar{x}_{n\gamma})\right)\nabla U_\gamma(\bar{x}_{n\gamma})\,dr\right|^2\right],\\
G_2(t)&=\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left|-\int_{n\gamma}^{t}\nabla^2U(\bar{x}_r)\left(\nabla U_{1,\gamma}(r,\bar{x}_{n\gamma})+\nabla U_{2,\gamma}(r,\bar{x}_{n\gamma})\right)\,dr\right|^2\right],\\
G_3(t)&=\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left|\sqrt{2}\int_{n\gamma}^{t}\left(\nabla^2U(\bar{x}_r)-\nabla^2U(\bar{x}_{n\gamma})\right)\,dw_r\right|^2\right],\\
G_4(t)&=\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left|\int_{n\gamma}^{t}\left(\vec{\Delta}(\nabla U)(\bar{x}_r)-\vec{\Delta}(\nabla U)(\bar{x}_{n\gamma})\right)\,dr\right|^2\right],\\
G_5(t)&=\mathbb{E}^{\mathscr{F}_{n\gamma}}\Big[\big(|\nabla^2U(\bar{x}_{n\gamma})||\nabla U(\bar{x}_{n\gamma})|^2\gamma^2+|\bar{x}_{n\gamma}||\nabla^2U(\bar{x}_{n\gamma})|^2|\nabla U(\bar{x}_{n\gamma})|^2\gamma^2\\
&\qquad+\gamma^{3/2}|\bar{x}_{n\gamma}||\vec{\Delta}(\nabla U)(\bar{x}_{n\gamma})|^2+\sqrt{2}\gamma|\nabla^2U(\bar{x}_{n\gamma})|^2|w_t-w_{n\gamma}|\big)^2\Big].
\end{aligned}$$

By using Cauchy-Schwarz inequality, Proposition 3.10, Remark 3.2 and Lemma 3.12, one obtains

$$\begin{aligned}
G_1(t)&\leq\gamma\int_{n\gamma}^{t}\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|(\nabla^2U(\bar{x}_r)-\nabla^2U(\bar{x}_{n\gamma}))\nabla U_\gamma(\bar{x}_{n\gamma})|^2\right]\,dr\\
&\leq C\gamma\int_{n\gamma}^{t}\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[(1+|\bar{x}_r|+|\bar{x}_{n\gamma}|)^{4\rho-4+4\alpha}|\bar{x}_r-\bar{x}_{n\gamma}|^2\right]\,dr\\
&\leq C\gamma\int_{n\gamma}^{t}\sqrt{\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[V_c(\bar{x}_r)+V_c(\bar{x}_{n\gamma})\right]\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\bar{x}_r-\bar{x}_{n\gamma}|^4\right]}\,dr\\
&\leq C\gamma^3V_c(\bar{x}_{n\gamma}).
\end{aligned} \tag{3.18}$$

Similarly, applying Cauchy-Schwarz inequality, Proposition 3.10 and Remark 3.2 yield

$$\begin{aligned}
G_2(t)&\leq\gamma\int_{n\gamma}^{t}\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\nabla^2U(\bar{x}_r)(\nabla U_{1,\gamma}(r,\bar{x}_{n\gamma})+\nabla U_{2,\gamma}(r,\bar{x}_{n\gamma}))|^2\right]\,dr\\
&\leq C\gamma\int_{n\gamma}^{t}\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[(1+|\bar{x}_r|)^{2\rho-2+2\alpha}|\nabla U_{1,\gamma}(r,\bar{x}_{n\gamma})+\nabla U_{2,\gamma}(r,\bar{x}_{n\gamma})|^2\right]\,dr\\
&\leq C\gamma\int_{n\gamma}^{t}\sqrt{\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[V_c(\bar{x}_r)\right]\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\nabla U_{1,\gamma}(r,\bar{x}_{n\gamma})|^4+|\nabla U_{2,\gamma}(r,\bar{x}_{n\gamma})|^4\right]}\,dr\\
&\leq C\gamma^3V_c(\bar{x}_{n\gamma}),
\end{aligned} \tag{3.19}$$

where the last inequality is obtained by applying Lemma 3.11. Moreover, one obtains

$$\begin{aligned}
G_3(t)&\leq C\int_{n\gamma}^{t}\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\nabla^2U(\bar{x}_r)-\nabla^2U(\bar{x}_{n\gamma})|^2\right]\,dr\\
&\leq C\int_{n\gamma}^{t}\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[(1+|\bar{x}_r|+|\bar{x}_{n\gamma}|)^{2\rho-4+2\alpha}|\bar{x}_r-\bar{x}_{n\gamma}|^2\right]\,dr\\
&\leq C\int_{n\gamma}^{t}\sqrt{\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[V_c(\bar{x}_r)+V_c(\bar{x}_{n\gamma})\right]\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\bar{x}_r-\bar{x}_{n\gamma}|^4\right]}\,dr\\
&\leq C\gamma^2V_c(\bar{x}_{n\gamma}).
\end{aligned}$$

Furthermore, using Cauchy-Schwarz inequality, Proposition 3.10, Lemma 3.12 and **B-2** yield

$$
\begin{aligned}
G_4(t) &\leq \gamma \int_{n\gamma}^{t} \mathbb{E}^{\mathscr{F}_{n\gamma}} \left[\left| \vec{\Delta}(\nabla U)(\bar{x}_r) - \vec{\Delta}(\nabla U)(\bar{x}_{n\gamma}) \right|^2 \right] dr \\
&\leq C\gamma \int_{n\gamma}^{t} \mathbb{E}^{\mathscr{F}_{n\gamma}} \left[(1 + |\bar{x}_r| + |\bar{x}_{n\gamma}|)^{2\rho-4} |\bar{x}_r - \bar{x}_{n\gamma}|^{2\alpha} \right] dr \\
&\leq C\gamma \int_{n\gamma}^{t} \sqrt{\mathbb{E}^{\mathscr{F}_{n\gamma}} \left[V_c(\bar{x}_r) + V_c(\bar{x}_{n\gamma}) \right] \mathbb{E}^{\mathscr{F}_{n\gamma}} \left[|\bar{x}_r - \bar{x}_{n\gamma}|^{4\alpha} \right]} \, dr \\
&\leq C\gamma^{2+\alpha} V_c(\bar{x}_{n\gamma}).
\end{aligned}
\tag{3.20}
$$

The estimate of $G_5(t)$ can be obtained by straightforwad calculations, which implies $G_5(t) \leq C\gamma^3 V_c(\bar{x}_{n\gamma})$. Therefore,

$$
\mathbb{E}^{\mathscr{F}_{n\gamma}} \left[|\nabla U(\bar{x}_t) - \nabla U(\bar{x}_{n\gamma}) - \nabla U_{1,\gamma}(t, \bar{x}_{n\gamma}) - \nabla U_{2,\gamma}(t, \bar{x}_{n\gamma})|^2 \right] \leq C\gamma^2 V_c(\bar{x}_{n\gamma}).
$$

$\square$

For any $x, \bar{x} \in \mathbb{R}^d$, denote by $M(x, \bar{x})$ a matrix whose (i,j)-th entry is given by

$$
M^{(i,j)}(x, \bar{x}) = \sum_{k=1}^{d} \frac{\partial^3 U(\bar{x})}{\partial x^{(i)} \partial x^{(j)} \partial x^{(k)}} (x^{(k)} - \bar{x}^{(k)}).
$$

One then obtains the following results.

Lemma 3.14. *Assume **B-2** holds. Then, there exists a constant $C > 0$ such that for any $x, \bar{x} \in \mathbb{R}^d$,*

$$
\left| \nabla^2 U(x) - \nabla^2 U(\bar{x}) - M(x, \bar{x}) \right| \leq dL(1 + |x| + |\bar{x}|)^{\rho-2} |x - \bar{x}|^{1+\alpha}.
$$

Proof. For $U : \mathbb{R}^d \to \mathbb{R}$, denote by $\nabla^2 U^{(i,j)}$ the (i,j)-th entry of the matrix $\nabla^2 U$. Denote by $g^{(i,j)}(t) = \nabla^2 U^{(i,j)}(tx + (1-t)\bar{x})$, for any $i, j = 1, \ldots, d$, $x, \bar{x} \in \mathbb{R}^d$ and $t \in [0,1]$. One then observes

$$
\begin{aligned}
&\left| g^{(i,j)}(1) - g^{(i,j)}(0) - M^{(i,j)}(x, \bar{x}) \right| \\
&= \left| \nabla^2 U^{(i,j)}(x) - \nabla^2 U^{(i,j)}(\bar{x}) - M^{(i,j)}(x, \bar{x}) \right| \\
&= \left| \left\langle \int_0^1 \nabla(\nabla^2 U^{(i,j)})(tx + (1-t)\bar{x}) \, dt, x - \bar{x} \right\rangle - \left\langle \nabla(\nabla^2 U^{(i,j)})(\bar{x}), x - \bar{x} \right\rangle \right| \\
&\leq \left| \int_0^1 \left(\nabla(\nabla^2 U^{(i,j)})(tx + (1-t)\bar{x}) - \nabla(\nabla^2 U^{(i,j)})(\bar{x}) \right) dt \right| |x - \bar{x}| \,.
\end{aligned}
$$

One obtains that by Cauchy-Schwarz inequality and **B-2**

$$
\begin{aligned}
&\left| \nabla^2 U(x) - \nabla^2 U(\bar{x}) - M(x, \bar{x}) \right| \\
&\leq \left| \nabla^2 U(x) - \nabla^2 U(\bar{x}) - M(x, \bar{x}) \right|_{\mathsf{F}} \\
&= \left(\sum_{i,j=1}^{d} \left(\nabla^2 U^{(i,j)}(x) - \nabla^2 U^{(i,j)}(\bar{x}) - M^{(i,j)}(x, \bar{x}) \right)^2 \right)^{1/2} \\
&= \left(\sum_{i,j=1}^{d} \left(\left\langle \int_0^1 \nabla(\nabla^2 U^{(i,j)})(tx + (1-t)\bar{x}) - \nabla(\nabla^2 U^{(i,j)})(\bar{x}) \, dt, x - \bar{x} \right\rangle \right)^2 \right)^{1/2} \\
&\leq \left(\int_0^1 \sum_{k=1}^{d} \left| \nabla^2 (\nabla U)^{(k)}(tx + (1-t)\bar{x}) - \nabla^2 (\nabla U)^{(k)}(\bar{x}) \right|_{\mathsf{F}}^2 dt \right)^{1/2} |x - \bar{x}| \\
&\leq dL(1 + |x| + |\bar{x}|)^{\rho-2} |x - \bar{x}|^{1+\alpha},
\end{aligned}
$$

which completes the proof. $\square$

Lemma 3.15. *Assume **B-1** and **B-2** are satisfied. Then, there exists a constant $C > 0$ such that for all $\gamma \in (0,1)$, $n \in \mathbb{N}$, and $t \in [n\gamma, (n+1)\gamma)$,*

$$\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left|\int_{n\gamma}^{t} M(\bar{x}_r, \bar{x}_{n\gamma})\, dw_r\right|^2\right] \leq C\gamma^2 V_c(\bar{x}_{n\gamma}).$$

Proof. By using conditional Itô's isometry and Lemma 3.12, one obtains,

$$\begin{aligned}
&\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left|\int_{n\gamma}^{t} M(\bar{x}_r, \bar{x}_{n\gamma})\, dw_r\right|^2\right] \\
&\leq C\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\int_{n\gamma}^{t} |M(\bar{x}_r, \bar{x}_{n\gamma})|^2\, dr\right] \\
&= C\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\int_{n\gamma}^{t} \left(\sum_{i,j=1}^{d}\left|\sum_{k=1}^{d} \frac{\partial^3 U(\bar{x}_{n\gamma})}{\partial x^{(i)}\partial x^{(j)}\partial x^{(k)}}(\bar{x}_r^{(k)} - \bar{x}_{n\gamma}^{(k)})\right|^2\right) dr\right] \\
&\leq C\int_{n\gamma}^{t} \mathbb{E}^{\mathscr{F}_{n\gamma}}\left[(1+|\bar{x}_{n\gamma}|)^{2(\rho-2+\alpha)}|\bar{x}_r - \bar{x}_{n\gamma}|^2\right] dr \\
&\leq C\gamma^2 V_c(\bar{x}_{n\gamma}).
\end{aligned}$$

$\square$

Lemma 3.16. *Assume **B-1** and **B-2** are satisfied. Then, there exists a constant $C > 0$ such that for all $\gamma \in (0,1)$, $n \in \mathbb{N}$, and $t \in [n\gamma, (n+1)\gamma)$,*

$$\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left\langle\int_{n\gamma}^{t} (\nabla U(x_r) - \nabla U(\bar{x}_r))\, dr, \int_{n\gamma}^{t} M(\bar{x}_r, \bar{x}_{n\gamma})\, dw_r\right\rangle\right] \leq C\gamma^3(V_c(\bar{x}_{n\gamma}) + V_c(x_{n\gamma})).$$

Proof. For any $t \in [n\gamma, (n+1)\gamma)$, one observes that

$$\begin{aligned}
&\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left\langle\int_{n\gamma}^{t} (\nabla U(x_r) - \nabla U(\bar{x}_r))\, dr, \int_{n\gamma}^{t} M(\bar{x}_r, \bar{x}_{n\gamma})\, dw_r\right\rangle\right] \\
&= \mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left\langle\int_{n\gamma}^{t} \{\nabla U(x_r) - \nabla U(x_{n\gamma}) - (\nabla U(\bar{x}_r) - \nabla U(\bar{x}_{n\gamma}))\right.\right. \\
&\quad \left.\left.-\sqrt{2}\int_{n\gamma}^{r} \nabla^2 U(x_{n\gamma})\, dw_s + \sqrt{2}\int_{n\gamma}^{r} \nabla^2 U(\bar{x}_{n\gamma})\, dw_s\right\} dr, \int_{n\gamma}^{t} M(\bar{x}_r, \bar{x}_{n\gamma})\, dw_r\right\rangle\right] \\
&\quad + \sqrt{2}\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left\langle\int_{n\gamma}^{t}\int_{n\gamma}^{r} (\nabla^2 U(x_{n\gamma}) - \nabla^2 U(\bar{x}_{n\gamma}))\, dw_s\, dr, \int_{n\gamma}^{t} M(\bar{x}_r, \bar{x}_{n\gamma})\, dw_r\right\rangle\right].
\end{aligned} \tag{3.21}$$

The second term in (3.21) can be rewritten as

$$\begin{aligned}
&\sqrt{2}\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left\langle\int_{n\gamma}^{t}\int_{n\gamma}^{r} (\nabla^2 U(x_{n\gamma}) - \nabla^2 U(\bar{x}_{n\gamma}))\, dw_s\, dr, \int_{n\gamma}^{t} M(\bar{x}_r, \bar{x}_{n\gamma})\, dw_r\right\rangle\right] \\
&= \sqrt{2}\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\sum_{i=1}^{d}\int_{n\gamma}^{t}\sum_{l=1}^{d}\int_{n\gamma}^{r} (\nabla^2 U^{(i,l)}(x_{n\gamma}) - \nabla^2 U^{(i,l)}(\bar{x}_{n\gamma}))\, dw_s^{(l)}\, dr\right. \\
&\quad \left.\times \sum_{j=1}^{d}\int_{n\gamma}^{t}\sum_{k=1}^{d} \frac{\partial^3 U(\bar{x}_{n\gamma})}{\partial x^{(i)}\partial x^{(j)}\partial x^{(k)}}\left(-\int_{n\gamma}^{r} \nabla\tilde{U}_\gamma^{(k)}(s, \bar{x}_{n\gamma})\, ds + \sqrt{2}\int_{n\gamma}^{r} dw_s^{(k)}\right) dw_r^{(j)}\right] \\
&\leq C\gamma^3(V_c(x_{n\gamma}) + V_c(\bar{x}_{n\gamma})).
\end{aligned}$$

where the last inequality holds due to Cauchy-Schwarz inequality, Lemma 3.11, Proposition

3.9, 3.10 and the fact that for any $i, l, j, k = 1, \dots, d$

$$\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\int_{n\gamma}^{t}\int_{n\gamma}^{r}(\nabla^2 U^{(i,l)}(x_{n\gamma}) - \nabla^2 U^{(i,l)}(\bar{x}_{n\gamma}))\, dw_s^{(l)}\, dr \right.$$
$$\left. \times \int_{n\gamma}^{t}\frac{\partial^3 U(\bar{x}_{n\gamma})}{\partial x^{(i)}\partial x^{(j)}\partial x^{(k)}}\int_{n\gamma}^{r}\sqrt{2}\, dw_s^{(k)}\, dw_r^{(j)}\right] = 0.$$

Then, to estimate the first term of (3.21), one applies Itô's formula to $\nabla U(x_r)$ and $\nabla U(\bar{x}_r)$ to obtain, almost surely

$$\begin{aligned}
&\nabla U(x_r) - \nabla U(x_{n\gamma}) - (\nabla U(\bar{x}_r) - \nabla U(\bar{x}_{n\gamma})) \\
&\quad - \sqrt{2}\int_{n\gamma}^{r}\nabla^2 U(x_{n\gamma})\, dw_s + \sqrt{2}\int_{n\gamma}^{r}\nabla^2 U(\bar{x}_{n\gamma})\, dw_s \\
&= -\int_{n\gamma}^{r}\left(\nabla^2 U(x_s)\nabla U(x_s) - \vec{\Delta}(\nabla U)(x_s)\right) ds \\
&\quad + \sqrt{2}\int_{n\gamma}^{r}(\nabla^2 U(x_s) - \nabla^2 U(x_{n\gamma}))\, dw_s \\
&\quad + \int_{n\gamma}^{r}\left(\nabla^2 U(\bar{x}_s)\nabla \tilde{U}_\gamma(s, \bar{x}_{n\gamma}) - \vec{\Delta}(\nabla U)(\bar{x}_s)\right) ds \\
&\quad - \sqrt{2}\int_{n\gamma}^{r}(\nabla^2 U(\bar{x}_s) - \nabla^2 U(\bar{x}_{n\gamma}))\, dw_s.
\end{aligned} \tag{3.22}$$

By using Cauchy-Schwarz inequality and Lemma 3.15, equation (3.21) yields

$$\begin{aligned}
&\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left\langle\int_{n\gamma}^{t}(\nabla U(x_r) - \nabla U(\bar{x}_r))\, dr, \int_{n\gamma}^{t}M(\bar{x}_r, \bar{x}_{n\gamma})\, dw_r\right\rangle\right] \\
&\leq \sqrt{C\gamma^2 V_c(\bar{x}_{n\gamma})}\left(\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\gamma\int_{n\gamma}^{t}|\nabla U(x_r) - \nabla U(x_{n\gamma}) - (\nabla U(\bar{x}_r) - \nabla U(\bar{x}_{n\gamma}))\right.\right. \\
&\quad \left.\left. -\sqrt{2}\int_{n\gamma}^{r}\nabla^2 U(x_{n\gamma})\, dw_s + \sqrt{2}\int_{n\gamma}^{r}\nabla^2 U(\bar{x}_{n\gamma})\, dw_s|^2\, dr\right]\right)^{1/2} \\
&\quad + C\gamma^3(V_c(\bar{x}_{n\gamma}) + V_c(x_{n\gamma})).
\end{aligned}$$

Then, by taking into consideration (3.22), and by applying Remark 3.2, Proposition 3.9 and 3.10, one obtains

$$\begin{aligned}
&\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left\langle\int_{n\gamma}^{t}(\nabla U(x_r) - \nabla U(\bar{x}_r))\, dr, \int_{n\gamma}^{t}M(\bar{x}_r, \bar{x}_{n\gamma})\, dw_r\right\rangle\right] \\
&\leq \sqrt{C\gamma^2 V_c(\bar{x}_{n\gamma})}\left(\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\gamma^2\int_{n\gamma}^{t}\int_{n\gamma}^{r}\left|\nabla^2 U(x_s)\nabla U(x_s) - \vec{\Delta}(\nabla U)(x_s)\right|^2 ds\, dr\right]\right)^{1/2} \\
&\quad + \sqrt{C\gamma^2 V_c(\bar{x}_{n\gamma})}\left(\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\gamma^2\int_{n\gamma}^{t}\int_{n\gamma}^{r}\left|\nabla^2 U(\bar{x}_s)\nabla \tilde{U}_\gamma(s, \bar{x}_{n\gamma}) - \vec{\Delta}(\nabla U)(\bar{x}_s)\right|^2 ds\, dr\right]\right)^{1/2} \\
&\quad + \sqrt{C\gamma^2 V_c(\bar{x}_{n\gamma})}\left(\gamma\int_{n\gamma}^{t}\int_{n\gamma}^{r}\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left|\nabla^2 U(x_s) - \nabla^2 U(x_{n\gamma})\right|^2\right] ds\, dr\right)^{1/2} \\
&\quad + \sqrt{C\gamma^2 V_c(\bar{x}_{n\gamma})}\left(\gamma\int_{n\gamma}^{t}\int_{n\gamma}^{r}\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left|\nabla^2 U(\bar{x}_s) - \nabla^2 U(\bar{x}_{n\gamma})\right|^2\right] ds\, dr\right)^{1/2} \\
&\quad + C\gamma^3(V_c(\bar{x}_{n\gamma}) + V_c(x_{n\gamma})) \\
&\leq \sqrt{C\gamma^2 V_c(\bar{x}_{n\gamma})}\left(\gamma\int_{n\gamma}^{t}\int_{n\gamma}^{r}\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[(1 + |x_s| + |x_{n\gamma}|)^{2\rho-2}|x_s - x_{n\gamma}|^2\right] ds\, dr\right)^{1/2}
\end{aligned}$$

$$+ \sqrt{C\gamma^2 V_c(\bar{x}_{n\gamma})} \left(\gamma \int_{n\gamma}^{t} \int_{n\gamma}^{r} \mathbb{E}^{\mathscr{F}_{n\gamma}} \left[(1 + |\bar{x}_s| + |\bar{x}_{n\gamma}|)^{2\rho-2} |\bar{x}_s - \bar{x}_{n\gamma}|^2 \right] ds\, dr \right)^{1/2}$$
$$+ C\gamma^3 (V_c(\bar{x}_{n\gamma}) + V_c(x_{n\gamma})).$$

Finally by using Cauchy-Schwarz inequality and Lemma 3.12, one obtains

$$\mathbb{E}^{\mathscr{F}_{n\gamma}} \left[\left\langle \int_{n\gamma}^{t} (\nabla U(x_r) - \nabla U(\bar{x}_r))\, dr, \int_{n\gamma}^{t} M(\bar{x}_r, \bar{x}_{n\gamma})\, dw_r \right\rangle \right]$$
$$\leq C\gamma^3 (V_c(\bar{x}_{n\gamma}) + V_c(x_{n\gamma})).$$

□

Proof of Theorem 3.4. For $t > 0$, consider the coupling

$$\begin{cases} x_t = x_0 - \int_0^t \nabla U(x_r) dr + \sqrt{2} w_t, \\ \bar{x}_t = \bar{x}_0 - \int_0^t \nabla \tilde{U}_\gamma(r, \bar{x}_{\lfloor r/\gamma \rfloor \gamma})\, dr + \sqrt{2} w_t, \end{cases}$$

where $-\nabla \tilde{U}_\gamma(r, \bar{x}_{\lfloor r/\gamma \rfloor \gamma})$ is defined in (3.16). Let $(x_0, \bar{x}_0)$ be distributed according to ζ_0, where $\zeta_0 = \pi \otimes \delta_x$ for all $x \in \mathbb{R}^d$. Define $e_t = x_t - \bar{x}_t$, for all $t \in [n\gamma, (n+1)\gamma)$, $n \in \mathbb{N}$. By Itô's formula, one obtains, almost surely,

$$|e_t|^2 = |e_{n\gamma}|^2 - 2 \int_{n\gamma}^{t} \left\langle e_s, \nabla U(x_s) - \nabla \tilde{U}_\gamma(s, \bar{x}_{n\gamma}) \right\rangle ds.$$

Then, taking the expectation and taking the derivative on both sides yield

$$\begin{aligned} \frac{d}{dt} \mathbb{E}\left[|e_t|^2\right] &= -2\mathbb{E}\left[\left\langle e_t, \nabla U(x_t) - \nabla \tilde{U}_\gamma(t, \bar{x}_{n\gamma}) \right\rangle \right] \\ &= 2\mathbb{E}\left[\langle e_t, -(\nabla U(x_t) - \nabla U(\bar{x}_t)) \rangle \right] \\ &\quad + 2\mathbb{E}\left[\langle e_t, -(\nabla U(\bar{x}_t) - \nabla U(\bar{x}_{n\gamma}) - \nabla U_{1,\gamma}(t, \bar{x}_{n\gamma}) - \nabla U_{2,\gamma}(t, \bar{x}_{n\gamma})) \rangle \right] \\ &\quad + 2\mathbb{E}\left[\langle e_t, -(\nabla U(\bar{x}_{n\gamma}) - \nabla U_\gamma(\bar{x}_{n\gamma})) \rangle \right], \end{aligned}$$

which implies by using **B-3** and $|a||b| \leq \varepsilon a^2 + (4\varepsilon)^{-1} b^2$, $\varepsilon > 0$,

$$\begin{aligned} \frac{d}{dt} \mathbb{E}\left[|e_t|^2\right] &\leq (2\varepsilon)^{-1} \gamma^3 \mathbb{E}\left[|\nabla U(\bar{x}_{n\gamma})|^5 \right] - 2(m - \varepsilon) \mathbb{E}\left[|e_t|^2\right] \\ &\quad + 2\mathbb{E}\left[\langle e_t, -(\nabla U(\bar{x}_t) - \nabla U(\bar{x}_{n\gamma}) - \nabla U_{1,\gamma}(t, \bar{x}_{n\gamma}) - \nabla U_{2,\gamma}(t, \bar{x}_{n\gamma})) \rangle \right]. \end{aligned} \tag{3.23}$$

By applying Itô's formula to $\nabla U(\bar{x}_t)$, and by calculating $\nabla U(\bar{x}_t) - \nabla U(\bar{x}_{n\gamma}) - \nabla U_{1,\gamma}(t, \bar{x}_{n\gamma}) - \nabla U_{2,\gamma}(t, \bar{x}_{n\gamma})$, one obtains (3.17). Substituting (3.17) into (3.23) gives

$$\begin{aligned} &\frac{d}{dt} \mathbb{E}\left[|e_t|^2\right] \\ &\leq (2\varepsilon)^{-1} \gamma^3 \mathbb{E}\left[|\nabla U(\bar{x}_{n\gamma})|^5 \right] - 2(m - \varepsilon) \mathbb{E}\left[|e_t|^2\right] \\ &\quad + 2\mathbb{E}\left[|e_t| \left| \int_{n\gamma}^{t} (\nabla^2 U(\bar{x}_r) - \nabla^2 U(\bar{x}_{n\gamma})) \nabla U_\gamma(\bar{x}_{n\gamma})\, dr \right| \right] \\ &\quad + 2\mathbb{E}\left[|e_t| \left| \int_{n\gamma}^{t} \nabla^2 U(\bar{x}_r) (\nabla U_{1,\gamma}(r, \bar{x}_{n\gamma}) + \nabla U_{2,\gamma}(r, \bar{x}_{n\gamma}))\, dr \right| \right] \\ &\quad + 2\sqrt{2}\mathbb{E}\left[\left\langle e_t, \left(- \int_{n\gamma}^{t} (\nabla^2 U(\bar{x}_r) - \nabla^2 U(\bar{x}_{n\gamma}))\, dw_r \right) \right\rangle \right] \\ &\quad + 2\mathbb{E}\left[|e_t| \left| \int_{n\gamma}^{t} (\vec{\Delta}(\nabla U)(\bar{x}_r) - \vec{\Delta}(\nabla U)(\bar{x}_{n\gamma}))\, dr \right| \right] \\ &\quad + 2\mathbb{E}\left[|e_t| (|\nabla^2 U(\bar{x}_{n\gamma})| |\nabla U(\bar{x}_{n\gamma})|^2 \gamma^2 + |\bar{x}_{n\gamma}| |\nabla^2 U(\bar{x}_{n\gamma})|^2 |\nabla U(\bar{x}_{n\gamma})|^2 \gamma^2 \right. \end{aligned}$$

$$+\gamma^{3/2}|\bar{x}_{n\gamma}||\vec{\Delta}(\nabla U)(\bar{x}_{n\gamma})|^2 + \sqrt{2}\gamma|\nabla^2 U(\bar{x}_{n\gamma})|^2(w_t - w_{n\gamma}))\Big].$$

By Young's inequality and Cauchy-Schwarz inequality,

$$\frac{d}{dt}\mathbb{E}\left[|e_t|^2\right] \leq J_1(t) + J_2(t), \tag{3.24}$$

where

$$J_1(t) = 2\sqrt{2}\mathbb{E}\left[\left\langle e_t, \left(-\int_{n\gamma}^t (\nabla^2 U(\bar{x}_r) - \nabla^2 U(\bar{x}_{n\gamma}))\, dw_r\right)\right\rangle\right],$$

and

$$\begin{aligned}
&J_2(t) \\
&= (2\varepsilon)^{-1}\gamma^3\mathbb{E}\left[|\nabla U(\bar{x}_{n\gamma})|^5\right] - 2(m - 5\varepsilon)\mathbb{E}\left[|e_t|^2\right] \\
&\quad + (2\varepsilon)^{-1}\gamma\mathbb{E}\left[\int_{n\gamma}^t |(\nabla^2 U(\bar{x}_r) - \nabla^2 U(\bar{x}_{n\gamma}))\nabla U_\gamma(\bar{x}_{n\gamma})|^2\, dr\right] \\
&\quad + (2\varepsilon)^{-1}\gamma\mathbb{E}\left[\int_{n\gamma}^t |\nabla^2 U(\bar{x}_r)(\nabla U_{1,\gamma}(r, \bar{x}_{n\gamma}) + \nabla U_{2,\gamma}(r, \bar{x}_{n\gamma}))|^2\, dr\right] \\
&\quad + (2\varepsilon)^{-1}\gamma\mathbb{E}\left[\int_{n\gamma}^t |\vec{\Delta}(\nabla U)(\bar{x}_r) - \vec{\Delta}(\nabla U)(\bar{x}_{n\gamma})|^2\, dr\right] \\
&\quad + 4(2\varepsilon)^{-1}\gamma^3\mathbb{E}\Big[|\nabla^2 U(\bar{x}_{n\gamma})|^2|\nabla U(\bar{x}_{n\gamma})|^4 + |\bar{x}_{n\gamma}|^2|\nabla^2 U(\bar{x}_{n\gamma})|^4|\nabla U(\bar{x}_{n\gamma})|^4 \\
&\qquad\qquad + |\bar{x}_{n\gamma}|^2|\vec{\Delta}(\nabla U)(\bar{x}_{n\gamma})|^4 + 2|\nabla^2 U(\bar{x}_{n\gamma})|^4\Big].
\end{aligned}$$

By taking $\varepsilon = \frac{m}{12}$, and by using the results form (3.18) - (3.20) in Lemma 3.13, one obtains

$$J_2(t) \leq C\gamma^{2+\alpha}\mathbb{E}\left[V_c(\bar{x}_{n\gamma})\right] - \frac{7}{6}m\mathbb{E}\left[|e_t|^2\right], \tag{3.25}$$

where $\alpha = (0, 1]$. Moreover, one can rewrite $J_1(t)$ as follows

$$\begin{aligned}
J_1(t) &= -2\sqrt{2}\mathbb{E}\left[\left\langle e_t, \int_{n\gamma}^t (\nabla^2 U(\bar{x}_r) - \nabla^2 U(\bar{x}_{n\gamma}) - M(\bar{x}_r, \bar{x}_{n\gamma}))\, dw_r\right\rangle\right] \\
&\quad - 2\sqrt{2}\mathbb{E}\left[\left\langle e_t - e_{n\gamma}, \int_{n\gamma}^t M(\bar{x}_r, \bar{x}_{n\gamma})\, dw_r\right\rangle\right] \\
&\quad - 2\sqrt{2}\mathbb{E}\left[\left\langle e_{n\gamma}, \int_{n\gamma}^t M(\bar{x}_r, \bar{x}_{n\gamma})\, dw_r\right\rangle\right],
\end{aligned}$$

which implies that, due to Young's inequality, Lemma 3.14, 3.12 and the fact that the last term above is zero,

$$\begin{aligned}
J_1(t) &\leq 2\varepsilon\mathbb{E}\left[|e_t|^2\right] + C\gamma^{2+\alpha}\mathbb{E}\left[V_c(\bar{x}_{n\gamma})\right] \\
&\quad + 2\sqrt{2}\mathbb{E}\left[\left\langle \int_{n\gamma}^t (\nabla U(x_r) - \nabla\tilde{U}_\gamma(r, \bar{x}_{n\gamma}))\, dr, \int_{n\gamma}^t M(\bar{x}_r, \bar{x}_{n\gamma})\, dw_r\right\rangle\right].
\end{aligned}$$

It can be further rewritten as

$$\begin{aligned}
J_1(t) &\leq 2\varepsilon\mathbb{E}\left[|e_t|^2\right] + C\gamma^{2+\alpha}\mathbb{E}\left[V_c(\bar{x}_{n\gamma})\right] \\
&\quad + 2\sqrt{2}\mathbb{E}\left[\left\langle \int_{n\gamma}^t (\nabla U(x_r) - \nabla U(\bar{x}_r))\, dr, \int_{n\gamma}^t M(\bar{x}_r, \bar{x}_{n\gamma})\, dw_r\right\rangle\right] \\
&\quad + 2\sqrt{2}\mathbb{E}\Bigg[\Bigg\langle \int_{n\gamma}^t (\nabla U(\bar{x}_r) - \nabla U(\bar{x}_{n\gamma}) - \nabla U_{1,\gamma}(r, \bar{x}_{n\gamma}) - \nabla U_{2,\gamma}(r, \bar{x}_{n\gamma}))\, dr, \\
&\qquad\qquad \int_{n\gamma}^t M(\bar{x}_r, \bar{x}_{n\gamma})\, dw_r\Bigg\rangle\Bigg]
\end{aligned}$$

$$+ 2\sqrt{2}\mathbb{E}\left[\left\langle \int_{n\gamma}^{t} (\nabla U(\bar{x}_{n\gamma}) - \nabla U_\gamma(\bar{x}_{n\gamma}))\, dr, \int_{n\gamma}^{t} M(\bar{x}_r, \bar{x}_{n\gamma})\, dw_r \right\rangle\right],$$

which, by using Cauchy-Schwarz inequality, Remark 3.2, Lemma 3.16, 3.13 and 3.15, yields

$$J_1(t) \leq 2\varepsilon\mathbb{E}\left[|e_t|^2\right] + C\gamma^{2+\alpha}\mathbb{E}\left[V_c(x_{n\gamma}) + V_c(\bar{x}_{n\gamma})\right] \tag{3.26}$$

Substituting (3.26) and (3.25) into (3.24) with $\varepsilon = \frac{m}{12}$, one obtains the following result,

$$\frac{d}{dt}\mathbb{E}\left[|e_t|^2\right] \leq -m\mathbb{E}\left[|e_t|^2\right] + C\gamma^{2+\alpha}\mathbb{E}\left[V_c(x_{n\gamma}) + V_c(\bar{x}_{n\gamma})\right].$$

The application of Gronwall's lemma yields

$$\mathbb{E}\left[|e_t|^2\right] \leq e^{-m(t-n\gamma)}\mathbb{E}\left[|e_{n\gamma}|^2\right] + C\gamma^{3+\alpha}\mathbb{E}\left[V_c(x_{n\gamma}) + V_c(\bar{x}_{n\gamma})\right].$$

Finally, by induction, Proposition 3.9 and 3.10, one obtains

$$\begin{aligned}
\mathbb{E}\left[|e_{(n+1)\gamma}|^2\right] &\leq e^{-m\gamma(n+1)}\mathbb{E}\left[|e_0|^2\right] \\
&\quad + C\gamma^{3+\alpha}\sum_{k=0}^{n}\mathbb{E}\left[V_c(\bar{x}_{k\gamma}) + V_c(x_{k\gamma})\right]e^{-m\gamma(n-k)} \\
&\leq e^{-m\gamma(n+1)}\mathbb{E}\left[|x_0 - \bar{x}_0|^2\right] + \frac{3bC}{7c^2m}e^{(\frac{7}{3}c^2+m)\gamma}\gamma^{2+\alpha} \\
&\quad + \frac{C}{m}\gamma^{2+\alpha}(\mathbb{E}\left[V_c(x_0)\right] + b_a)e^{m\gamma} \\
&\quad + C\gamma^{3+\alpha}\mathbb{E}\left[V_c(\bar{x}_0)\right]\sum_{k=0}^{n}e^{-\frac{7}{3}c^2\gamma k - m\gamma(n-k)},
\end{aligned}$$

where the last inequality holds by using $1 - e^{-m\gamma} \geq m\gamma e^{-m\gamma}$, and this indicates (see Appendix B.3 for a detailed proof)

$$\mathbb{E}\left[|e_{(n+1)\gamma}|^2\right] \leq e^{-m\gamma(n+1)}\mathbb{E}\left[|x_0 - \bar{x}_0|^2\right] + C\gamma^{2+\alpha}, \tag{3.27}$$

Note that $(x_0, \bar{x}_0)$ is distributed according to ζ_0, then (3.5) can be obtained by using Theorem 1 in [15] and the triangle inequality.

3.4.3 Proof of Theorem 3.5

By applying the following lemma, one can show that without using **B-3**, the rate of convergence in total variation norm is of order 1, which is properly stated in Theorem 3.5.

Lemma 3.17. *Assume **B-1** and **B-2** are satisfied. Let $p \in \mathbb{N}$ and ν_0 be a probability measure on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. There exists $C > 0$ such that for all $\gamma \in (0,1)$*

$$\mathrm{KL}(\nu_0 R_\gamma^p | \nu_0 P_{p\gamma}) \leq C\gamma^3 \int_{\mathbb{R}^d}\sum_{i=0}^{p-1}\left(\int_{\mathbb{R}^d} V_c(z)R_\gamma^i(y, dz)\right)\nu_0(dy).$$

Proof. Denote by μ_p^y and $\bar{\mu}_p^y$ the laws on $\mathcal{C}([0, p\gamma], \mathbb{R}^d)$ of the SDE (3.1) and of the linear interpolation (3.16) of the scheme both started at $y \in \mathbb{R}^d$. Denote by $(\mathscr{F}_t)_{t\geq 0}$ the filtration associated with $(w_t)_{t\geq 0}$, and by $(x_t, \bar{x}_t)_{t\geq 0}$ the unique strong solution of

$$\begin{cases} dx_t = -\nabla U(x_t)dt + \sqrt{2}dw_t, \\ d\bar{x}_t = -\nabla \tilde{U}_\gamma(t, \bar{x}_{\lfloor t/\gamma\rfloor\gamma})\, dt + \sqrt{2}w_t, \end{cases} \tag{3.28}$$

where $-\nabla \tilde{U}_\gamma(t, \bar{x}_{\lfloor t/\gamma\rfloor\gamma})$ is defined in (3.16). Then, by taking into consideration Definition 7 concerning diffusion type processes and Lemma 4.9 which refers to their representations in section 4.2 from [32], Theorem 7.19 in [32] can be applied to obtain the Radon-Nikodym derivative of μ_p^y w.r.t. $\bar{\mu}_p^y$, i.e.

$$\begin{aligned}\frac{d\mu_p^y}{d\bar{\mu}_p^y}((\bar{x}_t)_{t\in[0,p\gamma]}) = \exp\Bigg(\frac{1}{2}\int_0^{p\gamma}\Big\langle -\nabla U(\bar{x}_s) + \nabla\tilde{U}_\gamma(s, \bar{x}_{\lfloor s/\gamma\rfloor\gamma})), d\bar{x}_s\Big\rangle \\ -\frac{1}{4}\int_0^{p\gamma}\left(|\nabla U(\bar{x}_s)|^2 - |\nabla\tilde{U}_\gamma(s,\bar{x}_{\lfloor s/\gamma\rfloor\gamma})|^2\right)ds\Bigg).\end{aligned} \tag{3.29}$$

Note that the assumptions of Theorem 7.19 in [32] are satisfied due to proposition 3.9 and 3.10. By using (3.29), one obtains

$$\begin{aligned}\mathrm{KL}(\bar{\mu}_p^y|\mu_p^y) &= \mathbb{E}_y\left(-\log\left(\frac{d\mu_p^y}{d\bar{\mu}_p^y}((\bar{x}_t)_{t\in[0,p\gamma]})\right)\right) \\ &= \frac{1}{4}\int_0^{p\gamma}\mathbb{E}_y\left(\left|\nabla U(\bar{x}_s) - \nabla\tilde{U}_\gamma(s,\bar{x}_{\lfloor s/\gamma\rfloor\gamma})\right|^2\right)ds \\ &= \frac{1}{4}\sum_{i=0}^{p-1}\int_{i\gamma}^{(i+1)\gamma}\mathbb{E}_y\left(\left|\nabla U(\bar{x}_s) - \nabla\tilde{U}_\gamma(s,\bar{x}_{i\gamma})\right|^2\right)ds \\ &\le \frac{1}{2}\sum_{i=0}^{p-1}\int_{i\gamma}^{(i+1)\gamma}\mathbb{E}_y\Big(\mathbb{E}^{\mathscr{F}_{i\gamma}}\big(|\nabla U(\bar{x}_s) - \nabla U(\bar{x}_{i\gamma}) \\ &\qquad -\nabla U_{1,\gamma}(s,\bar{x}_{i\gamma}) - \nabla U_{2,\gamma}(s,\bar{x}_{i\gamma})|^2\big)\Big)\,ds \\ &\quad + \frac{1}{2}\sum_{i=0}^{p-1}\int_{i\gamma}^{(i+1)\gamma}\mathbb{E}_y\left(\mathbb{E}^{\mathscr{F}_{i\gamma}}\left(|\nabla U(\bar{x}_{i\gamma}) - \nabla U_\gamma(\bar{x}_{i\gamma})|^2\right)\right)ds \\ &\le C\gamma^3\sum_{i=0}^{p-1}\mathbb{E}_y\left(V_c(\bar{x}_{i\gamma})\right),\end{aligned}$$

where the last inequality holds due to Lemma 3.13. Then, by Theorem 4.1 in [29], it follows that

$$\mathrm{KL}(\delta_y R_\gamma^p|\delta_y P_{p\gamma}) \le \mathrm{KL}(\bar{\mu}_p^y|\mu_p^y) \le C\gamma^3\sum_{i=0}^{p-1}\mathbb{E}_y\left(V_c(\bar{x}_{i\gamma})\right).$$

Finally, applying the tower property yields the desired result,

$$\begin{aligned}\mathrm{KL}(\nu_0 R_\gamma^p|\nu_0 P_{p\gamma}) &\le C\gamma^3\sum_{i=0}^{p-1}\mathbb{E}\left(\mathbb{E}_y\left(V_c(\bar{x}_{i\gamma})\right)\right) \\ &= C\gamma^3\int_{\mathbb{R}^d}\sum_{i=0}^{p-1}\left(\int_{\mathbb{R}^d}V_c(z)R_\gamma^i(y,dz)\right)\nu_0(dy).\end{aligned}$$

□

Proof of Theorem 3.5. The proof follows along the same lines as the proof of Theorem 4 in [7], but for the completeness, the details are given below.

By Proposition 3.9, for all $n\in\mathbb{N}$ and $x\in\mathbb{R}^d$, we have

$$\begin{aligned}\|\delta_x R_\gamma^n - \pi\|_{V_c^{1/2}} &\le \|\delta_x P_{n\gamma} - \pi\|_{V_c^{1/2}} + \|\delta_x R_\gamma^n - \delta_x P_{n\gamma}\|_{V_c^{1/2}} \\ &\le C_{c/2}\rho_{c/2}^{n\gamma}V_c^{1/2}(x) + \|\delta_x R_\gamma^n - \delta_x P_{n\gamma}\|_{V_c^{1/2}}.\end{aligned}$$

Denote by $k_\gamma = \lceil\gamma^{-1}\rceil$, and by q_γ, r_γ the quotient and the remainder of the Euclidean division of n by k_γ, i.e. $n = q_\gamma k_\gamma + r_\gamma$. Then,

$$\|\delta_x R_\gamma^n - \delta_x P_{n\gamma}\|_{V_c^{1/2}} \le I_1 + I_2,$$

where

$$
\begin{aligned}
I_1 &= \|\delta_x R_\gamma^{q_\gamma k_\gamma} P_{r_\gamma \gamma} - \delta_x R_\gamma^n\|_{V_c^{1/2}} \\
I_2 &= \sum_{i=1}^{q_\gamma} \|\delta_x R_\gamma^{(i-1)k_\gamma} P_{(n-(i-1)k_\gamma)\gamma} - \delta_x R_\gamma^{ik_\gamma} P_{(n-ik_\gamma)\gamma}\|_{V_c^{1/2}} \\
&\le \sum_{i=1}^{q_\gamma} C_{c/2} \rho_{c/2}^{(n-ik_\gamma)\gamma} \|\delta_x R_\gamma^{(i-1)k_\gamma} P_{k_\gamma \gamma} - \delta_x R_\gamma^{ik_\gamma}\|_{V_c^{1/2}}
\end{aligned}
$$

By applying Lemma 24 in [14] to I_1, we have

$$
\begin{aligned}
\|\delta_x R_\gamma^{q_\gamma k_\gamma} P_{r_\gamma \gamma} - \delta_x R_\gamma^n\|_{V_c^{1/2}}^2 \le 2 \left(\delta_x R_\gamma^{q_\gamma k_\gamma} P_{r_\gamma \gamma}(V_c) + \delta_x R_\gamma^n(V_c)\right) \\
\times \mathrm{KL}(\delta_x R_\gamma^n | \delta_x R_\gamma^{q_\gamma k_\gamma} P_{r_\gamma \gamma}).
\end{aligned} \tag{3.30}
$$

Then, by Proposition 3.10 and Lemma 3.17, one obtains

$$
\begin{aligned}
\mathrm{KL}(\delta_x R_\gamma^n | \delta_x R_\gamma^{q_\gamma k_\gamma} P_{r_\gamma \gamma}) &\le C\gamma^3 \sum_{j=0}^{r_\gamma - 1} \int_{\mathbb{R}^d} V_c(z) \delta_x R_\gamma^{q_\gamma k_\gamma + j}(dz) \\
&\le C\gamma^3 (1 + \gamma^{-1}) \left(e^{-\frac{7}{3}c^2 q_\gamma k_\gamma \gamma} V_c(x) + \frac{3b}{7c^2} e^{\frac{7}{3}c^2 \gamma}\right),
\end{aligned} \tag{3.31}
$$

where the last inequality holds since $r_\gamma \le k_\gamma \le 1 + \gamma^{-1}$. Furthermore, by Proposition 3.9 and Proposition 3.10,

$$
\delta_x R_\gamma^{q_\gamma k_\gamma} P_{r_\gamma \gamma}(V_c) + \delta_x R_\gamma^n(V_c) \le 2 \left(e^{-\frac{7}{3}c^2 q_\gamma k_\gamma \gamma} V_c(x) + \frac{3b}{7c^2} e^{\frac{7}{3}c^2 \gamma} + b_c\right). \tag{3.32}
$$

Substituting (3.31) and (3.32) into (3.30) yields

$$
\begin{aligned}
I_1 &\le 2C^{1/2} \gamma^{3/2} (1 + \gamma^{-1})^{1/2} \left(e^{-\frac{7}{3}c^2 q_\gamma k_\gamma \gamma} V_c(x) + \frac{3b}{7c^2} e^{\frac{7}{3}c^2 \gamma} + b_c\right) \\
&\le C(\lambda^{n\gamma} V_c(x) + \gamma),
\end{aligned}
$$

where $\lambda \in (0, 1)$. By using similar arguments to I_2, one obtains (3.6).

3.5 Lipschitz case

In the context of a Lipschitz gradient, assume **B-3 - B-6** hold. Then, by **B-4** and **B-5**, one obtains, for any $x, y \in \mathbb{R}^d$

$$
|\nabla^2 U(x) y| \le L_1 |y|, \quad |\vec{\Delta}(\nabla U(x))| \le d L_2. \tag{3.33}
$$

One also notice that by [40, Theorem 2.1.12], under **B-3** and **B-4**, for all $x, y \in \mathbb{R}^d$,

$$
\langle x - y, \nabla U(x) - \nabla U(y) \rangle \ge \tilde{m} |x - y|^2 + \frac{1}{m + L_1} |\nabla U(x) - \nabla U(y)|^2, \tag{3.34}
$$

where we have set

$$
\tilde{m} = \frac{m L_1}{m + L_1}. \tag{3.35}
$$

The linear interpolation of the algorithm (3.8) becomes

$$
\tilde{x}_t = \tilde{x}_0 - \int_0^t \nabla \tilde{U}(s, \tilde{x}_{\lfloor s/\gamma \rfloor \gamma}) \, ds + \sqrt{2} w_t, \tag{3.36}
$$

for all $t \geq 0$, where

$$\nabla \tilde{U}(s, \tilde{x}_{\lfloor s/\gamma \rfloor \gamma}) = \nabla U(\tilde{x}_{\lfloor s/\gamma \rfloor \gamma}) + \nabla U_1(s, \tilde{x}_{\lfloor s/\gamma \rfloor \gamma}) + \nabla U_2(s, \tilde{x}_{\lfloor s/\gamma \rfloor \gamma}),$$

with

$$\begin{aligned}
&\nabla U_1(s, \tilde{x}_{\lfloor s/\gamma \rfloor \gamma}) \\
&= -\int_{\lfloor s/\gamma \rfloor \gamma}^{s} \left(\nabla^2 U(\tilde{x}_{\lfloor s/\gamma \rfloor \gamma}) \nabla U(\tilde{x}_{\lfloor s/\gamma \rfloor \gamma}) - \vec{\Delta}(\nabla U)(\tilde{x}_{\lfloor s/\gamma \rfloor \gamma}) \right) dr,
\end{aligned}$$

and

$$\nabla U_2(s, \tilde{x}_{\lfloor s/\gamma \rfloor \gamma}) = \sqrt{2} \int_{\lfloor s/\gamma \rfloor \gamma}^{s} \nabla^2 U(\tilde{x}_{\lfloor s/\gamma \rfloor \gamma})\, dw_r.$$

One notes that for any $n \in \mathbb{N}$, $\tilde{X}_n = \tilde{x}_{n\gamma}$.

3.5.1 Moment bounds

Proposition 3.18. *Assume **B-3** - **B-6** are satisfied. Let x^* be the unique minimizer of U. Then, for all $x \in \mathbb{R}^d$, $\gamma \in \left(0, \frac{1}{\tilde{m}} \wedge \frac{4}{5(m+L_1)} \wedge \frac{\sqrt{m}}{(m+L_1)\sqrt{L_1}} \wedge \sqrt[3]{\frac{1}{(m+L_1)L_1^2}}\right)$ and $n \in \mathbb{N}$,*

$$\mathbb{E}^{\mathscr{F}_0} |\tilde{x}_{(n+1)\gamma} - x^*|^2 \leq (1 - \tilde{m}\gamma)^{n+1} |\tilde{x}_0 - x^*|^2 + \frac{q_1}{\tilde{m}},$$

where $q_1 = \left(\frac{L_2^2}{2\tilde{m}} + \frac{3L_2^2}{2}\right) d^2 + (4L_1^2 + 4)d$ and $\tilde{m}$ is given in (3.35).

Proof. Denote by

$$\Delta_n = \tilde{x}_{n\gamma} - x^* - \nabla U(\tilde{x}_{n\gamma})\gamma + \frac{\gamma^2}{2} \left(\nabla^2 U(\tilde{x}_{n\gamma}) \nabla U(\tilde{x}_{n\gamma}) - \vec{\Delta}(\nabla U)(\tilde{x}_{n\gamma}) \right),$$

where x^* is the unique minimizer of U and one calculates

$$\begin{aligned}
&\mathbb{E}^{\mathscr{F}_{n\gamma}} |\tilde{x}_{(n+1)\gamma} - x^*|^2 \\
&= \mathbb{E}^{\mathscr{F}_{n\gamma}} \left| \Delta_n - \sqrt{2} \int_{n\gamma}^{(n+1)\gamma} \int_{n\gamma}^{r} \nabla^2 U(\tilde{x}_{n\gamma})\, dw_s\, dr + \sqrt{2} \int_{n\gamma}^{(n+1)\gamma} dw_r \right|^2 \\
&\leq |\Delta_n|^2 + 4\mathbb{E}^{\mathscr{F}_{n\gamma}} \left| \int_{n\gamma}^{(n+1)\gamma} \int_{n\gamma}^{r} \nabla^2 U(\tilde{x}_{n\gamma})\, dw_s\, dr \right|^2 + 4d\gamma \\
&\leq |\Delta_n|^2 + 4\gamma^3 L_1^2 d + 4\gamma d,
\end{aligned} \tag{3.37}$$

where the last inequality holds due to (3.33). Then, by using (3.33), (3.34) and the strong convexity condition of U, one obtains, for $\gamma \in \left(0, \frac{1}{\tilde{m}} \wedge \frac{4}{5(m+L_1)} \wedge \frac{\sqrt{m}}{(m+L_1)\sqrt{L_1}} \wedge \sqrt[3]{\frac{1}{(m+L_1)L_1^2}}\right)$ and $n \in \mathbb{N}$,

$$\begin{aligned}
|\Delta_n|^2 &= |\tilde{x}_{n\gamma} - x^*|^2 + \left| -\nabla U(\tilde{x}_{n\gamma})\gamma + \frac{\gamma^2}{2} \left(\nabla^2 U(\tilde{x}_{n\gamma}) \nabla U(\tilde{x}_{n\gamma}) - \vec{\Delta}(\nabla U)(\tilde{x}_{n\gamma}) \right) \right|^2 \\
&\quad + 2 \left\langle \tilde{x}_{n\gamma} - x^*, -\nabla U(\tilde{x}_{n\gamma})\gamma + \frac{\gamma^2}{2} \left(\nabla^2 U(\tilde{x}_{n\gamma}) \nabla U(\tilde{x}_{n\gamma}) - \vec{\Delta}(\nabla U)(\tilde{x}_{n\gamma}) \right) \right\rangle \\
&\leq (1 - 2\tilde{m}\gamma) |\tilde{x}_{n\gamma} - x^*|^2 - \frac{2\gamma}{m + L_1} |\nabla U(\tilde{x}_{n\gamma}) - \nabla U(x^*)|^2 \\
&\quad + \tilde{m}\gamma |\tilde{x}_{n\gamma} - x^*|^2 + \frac{\gamma^3}{4\tilde{m}} \left| \nabla^2 U(\tilde{x}_{n\gamma}) \nabla U(\tilde{x}_{n\gamma}) - \vec{\Delta}(\nabla U)(\tilde{x}_{n\gamma}) \right|^2 \\
&\quad + \gamma^2 |\nabla U(\tilde{x}_{n\gamma}) - \nabla U(x^*)|^2 + \frac{\gamma^4}{4} \left| \nabla^2 U(\tilde{x}_{n\gamma}) \nabla U(\tilde{x}_{n\gamma}) - \vec{\Delta}(\nabla U)(\tilde{x}_{n\gamma}) \right|^2
\end{aligned}$$

$$+ \gamma^3 \left\langle \nabla U(\tilde{x}_{n\gamma}), \vec{\Delta}(\nabla U)(\tilde{x}_{n\gamma}) \right\rangle,$$

which by using $(a+b)^2 \le 2a^2+2b^2$ and $2ab \le a^2+b^2$ for $a,b \ge 0$ yield

$$\begin{aligned}
|\Delta_n|^2 &\le (1-\tilde{m}\gamma)|\tilde{x}_{n\gamma}-x^*|^2 \\
&\quad + \left(-\frac{2\gamma}{m+L_1} + \frac{5\gamma^2}{4} + \frac{\gamma^3 L_1^2}{2\tilde{m}} + \frac{\gamma^4 L_1^2}{2} \right) |\nabla U(\tilde{x}_{n\gamma}) - \nabla U(x^*)|^2 \\
&\quad + \frac{\gamma^3 L_2^2}{2\tilde{m}} d^2 + \frac{\gamma^4 L_2^2}{2} d^2 + \gamma^4 L_2^2 d^2 \\
&\le (1-\tilde{m}\gamma)|\tilde{x}_{n\gamma}-x^*|^2 + \gamma^3 \left(\frac{L_2^2}{2\tilde{m}} + \frac{3L_2^2}{2} \right) d^2,
\end{aligned} \tag{3.38}$$

where $\tilde{m}$ is defined in (3.35). Substituting the above upper bound into (3.37) yields

$$\mathbb{E}^{\mathscr{F}_{n\gamma}} |\tilde{x}_{(n+1)\gamma} - x^*|^2 \le (1-\tilde{m}\gamma)|\tilde{x}_{n\gamma}-x^*|^2 + \gamma q_1,$$

where $q_1 = \left(\frac{L_2^2}{2\tilde{m}} + \frac{3L_2^2}{2} \right) d^2 + (4L_1^2+4)d$, and the result can be obtained by induction. $\square$

Proposition 3.19. *Assume **B-3** - **B-6** are satisfied. Let x^* be the unique minimizer of U. Then, for all $x \in \mathbb{R}^d$, $\gamma \in \left(0, \frac{1}{\tilde{m}} \wedge \frac{4}{5(m+L_1)} \wedge \frac{\sqrt{m}}{(m+L_1)\sqrt{L_1}} \wedge \sqrt[3]{\frac{1}{(m+L_1)L_1^2}}\right)$,*

$$\mathbb{E}^{\mathscr{F}_0} |\tilde{x}_{(n+1)\gamma} - x^*|^4 \le \left(1 - \frac{\tilde{m}\gamma}{8}\right)^{n+1} |\tilde{x}_0 - x^*|^4 + \frac{8q_2}{\tilde{m}},$$

where $q_2 = \left(2 + \frac{8}{\tilde{m}\gamma}\right) \left(\frac{L_2^2}{2\tilde{m}} + \frac{3L_2^2}{2} \right)^2 d^4 + 32\gamma \left(1 + \frac{42}{\tilde{m}}\right)(L_1^4+3)d^2$ and $\tilde{m}$ is given in (3.35).

Proof. Denote by

$$\Delta_n = \tilde{x}_{n\gamma} - x^* - \nabla U(\tilde{x}_{n\gamma})\gamma + \frac{\gamma^2}{2} \left(\nabla^2 U(\tilde{x}_{n\gamma}) \nabla U(\tilde{x}_{n\gamma}) - \vec{\Delta}(\nabla U)(\tilde{x}_{n\gamma}) \right),$$

and

$$\tilde{\Delta}_{n+1} = -\sqrt{2} \int_{n\gamma}^{(n+1)\gamma} \int_{n\gamma}^{r} \nabla^2 U(\tilde{x}_{n\gamma})\, dw_s\, dr + \sqrt{2} \int_{n\gamma}^{(n+1)\gamma} dw_r.$$

One obtains by using Jensen's inequality

$$\begin{aligned}
&\mathbb{E}^{\mathscr{F}_{n\gamma}} |\tilde{x}_{(n+1)\gamma} - x^*|^4 && (3.39) \\
&= \mathbb{E}^{\mathscr{F}_{n\gamma}} \left| \Delta_n + \tilde{\Delta}_{n+1} \right|^4 \\
&= \mathbb{E}^{\mathscr{F}_{n\gamma}} \left(|\Delta_n|^2 + 2\left\langle \Delta_n, \tilde{\Delta}_{n+1} \right\rangle + |\tilde{\Delta}_{n+1}|^2 \right)^2 \\
&= \mathbb{E}^{\mathscr{F}_{n\gamma}} \Big(|\Delta_n|^4 + 4\left\langle \Delta_n, \tilde{\Delta}_{n+1} \right\rangle |\Delta_n|^2 + 2|\Delta_n|^2 |\tilde{\Delta}_{n+1}|^2 + 4\left| \left\langle \Delta_n, \tilde{\Delta}_{n+1} \right\rangle \right|^2 && (3.40) \\
&\qquad\qquad + 4\left\langle \Delta_n, \tilde{\Delta}_{n+1} \right\rangle |\tilde{\Delta}_{n+1}|^2 + |\tilde{\Delta}_{n+1}|^4 \Big) \\
&\le |\Delta_n|^4 + 6|\Delta_n|^2 \mathbb{E}^{\mathscr{F}_{n\gamma}} |\tilde{\Delta}_{n+1}|^2 + \mathbb{E}^{\mathscr{F}_{n\gamma}} |\tilde{\Delta}_{n+1}|^4 + 4|\Delta_n| \mathbb{E}^{\mathscr{F}_{n\gamma}} |\tilde{\Delta}_{n+1}|^3 \\
&\le \left(1 + \frac{\tilde{m}\gamma}{2}\right) |\Delta_n|^4 + \frac{36}{\tilde{m}\gamma} \left(\mathbb{E}^{\mathscr{F}_{n\gamma}} |\tilde{\Delta}_{n+1}|^2 \right)^2 + \left(1 + \frac{6}{\tilde{m}\gamma}\right) \mathbb{E}^{\mathscr{F}_{n\gamma}} |\tilde{\Delta}_{n+1}|^4 \\
&\le \left(1 + \frac{\tilde{m}\gamma}{2}\right) |\Delta_n|^4 + 32\gamma \left(1 + \frac{42}{\tilde{m}}\right) (L_1^4 + 3) d^2. && (3.41)
\end{aligned}$$

Then, by using (3.38) and the inequality $(a+b)^2 \le (1+\epsilon)a^2 + (1+\epsilon^{-1})b^2$, for any $a, b \ge 0$,

$\epsilon > 0$, one obtains, for $\gamma \in \left(0, \frac{1}{\tilde{m}} \wedge \frac{4}{5(m+L_1)} \wedge \frac{\sqrt{m}}{(m+L_1)\sqrt{L_1}} \wedge \sqrt[3]{\frac{1}{(m+L_1)L_1^2}}\right)$,

$$
\begin{aligned}
&\mathbb{E}^{\mathscr{F}_{n\gamma}} |\tilde{x}_{(n+1)\gamma} - x^*|^4 \\
&\leq \left(1 + \frac{\tilde{m}\gamma}{2}\right)\left((1-\tilde{m}\gamma)|\tilde{x}_{n\gamma} - x^*|^2 + \gamma^3\left(\frac{L_2^2}{2\tilde{m}} + \frac{3L_2^2}{2}\right)d^2\right)^2 \\
&\quad + 32\gamma\left(1 + \frac{42}{\tilde{m}}\right)(L_1^4 + 3)d^2 \\
&\leq \left(1 + \frac{\tilde{m}\gamma}{2}\right)\left(1 + \frac{\tilde{m}\gamma}{4}\right)(1-\tilde{m}\gamma)|\tilde{x}_{n\gamma} - x^*|^4 \\
&\quad + \left(1 + \frac{\tilde{m}\gamma}{2}\right)\left(1 + \frac{4}{\tilde{m}\gamma}\right)\gamma^6\left(\frac{L_2^2}{2\tilde{m}} + \frac{3L_2^2}{2}\right)^2 d^4 \\
&\quad + 32\gamma\left(1 + \frac{42}{\tilde{m}}\right)(L_1^4 + 3)d^2 \\
&\leq \left(1 - \frac{\tilde{m}\gamma}{8}\right)|\tilde{x}_{n\gamma} - x^*|^4 + \gamma^6\left(2 + \frac{8}{\tilde{m}\gamma}\right)\left(\frac{L_2^2}{2\tilde{m}} + \frac{3L_2^2}{2}\right)^2 d^4 \\
&\quad + 32\gamma\left(1 + \frac{42}{\tilde{m}}\right)(L_1^4 + 3)d^2,
\end{aligned}
$$

which implies

$$
\mathbb{E}^{\mathscr{F}_{n\gamma}} |\tilde{x}_{(n+1)\gamma} - x^*|^4 \leq \left(1 - \frac{\tilde{m}\gamma}{8}\right)|\tilde{x}_{n\gamma} - x^*|^4 + \gamma q_2,
$$

where $q_2 = \left(2 + \frac{8}{\tilde{m}}\right)\left(\frac{L_2^2}{2\tilde{m}} + \frac{3L_2^2}{2}\right)^2 d^4 + 32\left(1 + \frac{42}{\tilde{m}}\right)(L_1^4 + 3)d^2$. The desired result follows by induction. □

3.5.2 Proof of Theorem 3.6

The explicit constants for the second and the fourth moments are obtained, then by using the following lemmas, one can show the rate of convergence in Wasserstein distance.

Lemma 3.20. *Let $\gamma \in \left(0, \frac{1}{\tilde{m}} \wedge \frac{4}{5(m+L_1)} \wedge \frac{\sqrt{m}}{(m+L_1)\sqrt{L_1}} \wedge \sqrt[3]{\frac{1}{(m+L_1)L_1^2}}\right)$. Assume* ***B-3*** *-* ***B-6*** *are satisfied. Then, for all $n \in \mathbb{N}$, and $t \in [n\gamma, (n+1)\gamma)$,*

$$
\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\nabla U_1(t, \tilde{x}_{n\gamma})|^2\right] \leq 2\gamma^2(L_1^4|\tilde{x}_{n\gamma} - x^*|^2 + d^2L_2^2),
$$

$$
\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\nabla U_1(t, \tilde{x}_{n\gamma})|^4\right] \leq 8\gamma^4(L_1^8|\tilde{x}_{n\gamma} - x^*|^4 + d^4L_2^4),
$$

$$
\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\nabla U_2(t, \tilde{x}_{n\gamma})|^2\right] \leq 2\gamma d L_1^2, \quad \mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\nabla U_2(t, \tilde{x}_{n\gamma})|^4\right] \leq 12L_1^4d^2\gamma^2.
$$

Proof. The proof is straightforward by using (3.33). □

Lemma 3.21. *Let $\gamma \in \left(0, \frac{1}{\tilde{m}} \wedge \frac{4}{5(m+L_1)} \wedge \frac{\sqrt{m}}{(m+L_1)\sqrt{L_1}} \wedge \sqrt[3]{\frac{1}{(m+L_1)L_1^2}}\right)$. Assume* ***B-3*** *-* ***B-6*** *are satisfied. Then, for all $n \in \mathbb{N}$, and $t \in [n\gamma, (n+1)\gamma)$,*

$$
\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\tilde{x}_t - \tilde{x}_{n\gamma}|^2\right] \leq \gamma(c_1|\tilde{x}_{n\gamma} - x^*|^2 + c_2),
$$

where $c_1 = \frac{5L_1^2}{4} + \frac{L_1^4}{2}$ and $c_2 = \frac{3L_2^2}{2}d^2 + 4L_1^2d + 4d$,

$$
\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\tilde{x}_t - \tilde{x}_{n\gamma}|^4\right] \leq \gamma^2(c_3|\tilde{x}_{n\gamma} - x^*|^4 + c_4),
$$

where $c_3 = 12\left(\frac{25}{16}L_1^4 + \frac{L_1^8}{4}\right)$ and $c_4 = \frac{54L_2^4}{4}d^4 + 416(L_1^4 + 3)d^2$, and

$$
\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|x_t - x_{n\gamma}|^2\right] \leq 2\gamma^2L_1^2|x_{n\gamma} - x^*|^2 + 4\gamma^3L_1^2d + 4\gamma d.
$$

Proof. One observes that

$$
\begin{aligned}
&\mathbb{E}^{\mathscr{F}_{n\gamma}}|\tilde{x}_t-\tilde{x}_{n\gamma}|^2\\
&=\left|-\nabla U(\tilde{x}_{n\gamma})(t-n\gamma)+\frac{(t-n\gamma)^2}{2}\left(\nabla^2U(\tilde{x}_{n\gamma})\nabla U(\tilde{x}_{n\gamma})-\vec{\Delta}(\nabla U)(\tilde{x}_{n\gamma})\right)\right|^2\\
&\quad+4\mathbb{E}^{\mathscr{F}_{n\gamma}}\left|\int_{n\gamma}^t\int_{n\gamma}^r\nabla^2U(\tilde{x}_{n\gamma})\,dw_s\,dr\right|^2+4d(t-n\gamma)\\
&\le|\nabla U(\tilde{x}_{n\gamma})-\nabla U(x^*)|^2\gamma^2+\frac{\gamma^4}{4}|\nabla^2U(\tilde{x}_{n\gamma})\nabla U(\tilde{x}_{n\gamma})-\vec{\Delta}(\nabla U)|^2\\
&\quad+\gamma^3\left|\left\langle\nabla U,\vec{\Delta}(\nabla U)\right\rangle\right|+4\gamma^3dL_1^2+4d\gamma\\
&\le\left(\frac{5\gamma^2}{4}+\frac{\gamma^4L_1^2}{2}\right)|\nabla U(\tilde{x}_{n\gamma})-\nabla U(x^*)|^2+\frac{\gamma^4L_2^2}{2}d^2+\gamma^4L_2^2d^2+4\gamma^3L_1^2d+4\gamma d \qquad (3.42)\\
&\le\gamma(c_1|\tilde{x}_{n\gamma}-x^*|^2+c_2),
\end{aligned}
$$

where $c_1=\frac{5L_1^2}{4}+\frac{L_1^4}{2}$ and $c_2=\frac{3L_2^2}{2}d^2+4L_1^2d+4d$. Then, denote by

$$
\bar{\Delta}_n=-\nabla U(\tilde{x}_{n\gamma})(t-n\gamma)+\frac{(t-n\gamma)^2}{2}\left(\nabla^2U(\tilde{x}_{n\gamma})\nabla U(\tilde{x}_{n\gamma})-\vec{\Delta}(\nabla U)(\tilde{x}_{n\gamma})\right)
$$

and recall

$$
\tilde{\Delta}_t=-\sqrt{2}\int_{n\gamma}^t\int_{n\gamma}^r\nabla^2U(\tilde{x}_{n\gamma})\,dw_s\,dr+\sqrt{2}\int_{n\gamma}^t dw_r.
$$

Notice that $|\bar{\Delta}_n|^2\le\gamma((\frac{5L_1^2}{4}+\frac{L_1^4}{2})|\tilde{x}_{n\gamma}-x^*|^2+\frac{3L_2^2}{2}d^2)$ by equation (3.42), and then one calculates

$$
\begin{aligned}
&\mathbb{E}^{\mathscr{F}_{n\gamma}}|\tilde{x}_t-\tilde{x}_{n\gamma}|^4\\
&=\mathbb{E}^{\mathscr{F}_{n\gamma}}|\bar{\Delta}_n+\tilde{\Delta}_t|^4\\
&=\mathbb{E}^{\mathscr{F}_{n\gamma}}\left(|\bar{\Delta}_n|^2+2\left\langle\bar{\Delta}_n,\tilde{\Delta}_t\right\rangle+|\tilde{\Delta}_t|^2\right)^2\\
&\le|\bar{\Delta}_n|^4+6|\bar{\Delta}_n|^2\mathbb{E}^{\mathscr{F}_{n\gamma}}|\tilde{\Delta}_t|^2+\mathbb{E}^{\mathscr{F}_{n\gamma}}|\tilde{\Delta}_t|^4+4|\bar{\Delta}_n|\mathbb{E}^{\mathscr{F}_{n\gamma}}|\tilde{\Delta}_t|^3\\
&\le3|\bar{\Delta}_n|^4+13\mathbb{E}^{\mathscr{F}_{n\gamma}}|\tilde{\Delta}_t|^4\\
&\le12\gamma^2\left(\frac{25}{16}L_1^4+\frac{L_1^8}{4}\right)|\tilde{x}_{n\gamma}-x^*|^4+\gamma^2\frac{54L_2^4}{4}d^4+416\gamma^2(L_1^4+3)d^2\\
&\le\gamma^2(c_3|\tilde{x}_{n\gamma}-x^*|^4+c_4),
\end{aligned}
$$

where $c_3=12\left(\frac{25}{16}L_1^4+\frac{L_1^8}{4}\right)$ and $c_4=\frac{54L_2^4}{4}d^4+416(L_1^4+3)d^2$. As for the third result, consider

$$
\begin{aligned}
\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|x_t-x_{n\gamma}|^2\right]&=\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left|-\int_{n\gamma}^t\nabla U(x_r)\,dr+\sqrt{2}\int_{n\gamma}^t dw_r\right|^2\right]\\
&\le2\gamma L_1^2\int_{n\gamma}^t\mathbb{E}^{\mathscr{F}_{n\gamma}}|x_r-x^*|^2\,dr+4\gamma d\\
&\le2\gamma^2L_1^2|x_{n\gamma}-x^*|^2+4\gamma^3L_1^2d+4\gamma d,
\end{aligned}
$$

where the last inequality holds by using Proposition 1 in [15]. ☐

Lemma 3.22. *Let $\gamma\in\left(0,\frac{1}{m}\wedge\frac{4}{5(m+L_1)}\wedge\frac{\sqrt{m}}{(m+L_1)\sqrt{L_1}}\wedge\sqrt[3]{\frac{1}{(m+L_1)L_1^2}}\right)$. Assume **B-3** - **B-6** are satisfied. Then, for all $n\in\mathbb{N}$, and $t\in[n\gamma,(n+1)\gamma)$,*

$$
\begin{aligned}
&\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\nabla U(\tilde{x}_t)-\nabla U(\tilde{x}_{n\gamma})-\nabla U_1(t,\tilde{x}_{n\gamma})-\nabla U_2(t,\tilde{x}_{n\gamma})|^2\right]\\
&\le\gamma^2(c_5|\tilde{x}_{n\gamma}-x^*|^4+c_6|\tilde{x}_{n\gamma}-x^*|^2+c_7),
\end{aligned}
$$

where c_5, c_6 and c_7 are given explicitly in the proof.

Proof. For any $t \in [n\gamma, (n+1)\gamma)$, applying Itô's formula to $\nabla U(\tilde{x}_t) - \nabla U(\tilde{x}_{n\gamma})$ gives, almost surely

$$\begin{aligned}
&\nabla U(\tilde{x}_t) - \nabla U(\tilde{x}_{n\gamma}) - \nabla U_1(t, \tilde{x}_{n\gamma}) - \nabla U_2(t, \tilde{x}_{n\gamma}) \\
&= -\int_{n\gamma}^{t} \left(\nabla^2 U(\tilde{x}_r) - \nabla^2 U(\tilde{x}_{n\gamma})\right) \nabla U(\tilde{x}_{n\gamma})\, dr \\
&\quad - \int_{n\gamma}^{t} \nabla^2 U(\tilde{x}_r) \left(\nabla U_1(r, \tilde{x}_{n\gamma}) + \nabla U_2(r, \tilde{x}_{n\gamma})\right)\, dr \\
&\quad + \sqrt{2}\int_{n\gamma}^{t} \left(\nabla^2 U(\tilde{x}_r) - \nabla^2 U(\tilde{x}_{n\gamma})\right)\, dw_r \\
&\quad + \int_{n\gamma}^{t} \left(\vec{\Delta}(\nabla U)(\tilde{x}_r) - \vec{\Delta}(\nabla U)(\tilde{x}_{n\gamma})\right)\, dr
\end{aligned} \tag{3.43}$$

Then, squaring both sides and taking conditional expectation gives

$$\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left|\nabla U(\tilde{x}_t) - \nabla U(\tilde{x}_{n\gamma}) - \nabla U_1(t, \tilde{x}_{n\gamma}) - \nabla U_2(t, \tilde{x}_{n\gamma})\right|^2\right] \le 4\sum_{i=1}^{4} \bar{G}_i(t). \tag{3.44}$$

By using Cauchy-Schwarz inequality, **B-4**, **B-5** and Lemma 3.21, one obtains

$$\begin{aligned}
\bar{G}_1(t) &\le \gamma \int_{n\gamma}^{t} \mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|(\nabla^2 U(\tilde{x}_r) - \nabla^2 U(\tilde{x}_{n\gamma}))\nabla U(\tilde{x}_{n\gamma})|^2\right]\, dr \\
&\le \gamma L_1^2 L_2^2 |\tilde{x}_{n\gamma} - x^*|^2 \int_{n\gamma}^{t} \mathbb{E}^{\mathscr{F}_{n\gamma}} |\tilde{x}_r - \tilde{x}_{n\gamma}|^2\, dr \\
&\le \gamma^3 (c_1 L_1^2 L_2^2 |\tilde{x}_{n\gamma} - x^*|^4 + c_2 L_1^2 L_2^2 |\tilde{x}_{n\gamma} - x^*|^2).
\end{aligned}$$

Similarly, by Cauchy-Schwarz inequality, (3.33) and Lemma 3.20, we have

$$\begin{aligned}
\bar{G}_2(t) &\le \gamma \int_{n\gamma}^{t} \mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\nabla^2 U(\tilde{x}_r)(\nabla U_1(r, \tilde{x}_{n\gamma}) + \nabla U_2(r, \tilde{x}_{n\gamma}))|^2\right]\, dr \\
&\le 2\gamma L_1^2 \int_{n\gamma}^{t} \mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\nabla U_1(r, \tilde{x}_{n\gamma})|^2 + |\nabla U_2(r, \tilde{x}_{n\gamma})|^2\right]\, dr \\
&\le 2\gamma^2 L_1^2 (2\gamma^2 (L_1^4 |\tilde{x}_{n\gamma} - x^*|^2 + d^2 L_2^2) + 2\gamma d L_1^2) \\
&\le \gamma^3 (4\gamma L_1^6 |\tilde{x}_{n\gamma} - x^*|^2 + 4\gamma L_1^2 L_2^2 d^2 + 4 d L_1^4).
\end{aligned}$$

Moreover, applying Cauchy-Schwarz inequality, **B-5** and Lemma 3.21 yields

$$\begin{aligned}
\bar{G}_3(t) &= 2\int_{n\gamma}^{t} \mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\nabla^2 U(\tilde{x}_r) - \nabla^2 U(\tilde{x}_{n\gamma})|_{\mathsf{F}}^2\right]\, dr \\
&\le 2 d L_2^2 \int_{n\gamma}^{t} \mathbb{E}^{\mathscr{F}_{n\gamma}} |\tilde{x}_r - \tilde{x}_{n\gamma}|^2\, dr \\
&\le \gamma^2 (2 L_2^2 d c_1 |\tilde{x}_{n\gamma} - x^*|^2 + 2 L_2^2 d c_2).
\end{aligned}$$

Furthermore, one obtains by using Cauchy-Schwarz inequality, **B-6** and Lemma 3.21

$$\begin{aligned}
\bar{G}_4(t) &\le \gamma \int_{n\gamma}^{t} \mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left|\vec{\Delta}(\nabla U)(\tilde{x}_r) - \vec{\Delta}(\nabla U)(\tilde{x}_{n\gamma})\right|^2\right]\, dr \\
&\le d^{3/2} L \gamma \int_{n\gamma}^{t} \mathbb{E}^{\mathscr{F}_{n\gamma}} |\tilde{x}_r - \tilde{x}_{n\gamma}|^2\, dr \\
&\le \gamma^3 (d^{3/2} L c_1 |\tilde{x}_{n\gamma} - x^*|^2 + d^{3/2} L c_2).
\end{aligned}$$

The proof completes by substituting all the estimates above into (3.44), i.e.

$$\begin{aligned}&\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[|\nabla U(\tilde{x}_t)-\nabla U(\tilde{x}_{n\gamma})-\nabla U_1(t,\tilde{x}_{n\gamma})-\nabla U_2(t,\tilde{x}_{n\gamma})|^2\right]\\&\le \gamma^2(c_5|\tilde{x}_{n\gamma}-x^*|^4+c_6|\tilde{x}_{n\gamma}-x^*|^2+c_7),\end{aligned}$$

where $c_5=4c_1L_1^2L_2^2$, $c_6=4(L_1^2L_2^2c_2+Lc_1d^{3/2}+2L_2^2c_1d+4L_1^6)$ and $c_7=4(Lc_2d^{3/2}+2L_2^2c_2d+4L_1^2L_2^2d^2+4L_1^4d)$. □

Lemma 3.23. *Let* $\gamma\in\left(0,\frac{1}{m}\wedge\frac{4}{5(m+L_1)}\wedge\frac{\sqrt{m}}{(m+L_1)\sqrt{L_1}}\wedge\sqrt[3]{\frac{1}{(m+L_1)L_1^2}}\right)$. *Assume* ***B-3*** *-* ***B-6*** *are satisfied. Then, for all* $n\in\mathbb{N}$, *and* $t\in[n\gamma,(n+1)\gamma)$,

$$\begin{aligned}&\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left\langle\int_{n\gamma}^t(\nabla U(x_r)-\nabla U(\tilde{x}_r))\,dr,\int_{n\gamma}^t(\nabla^2U(\tilde{x}_r)-\nabla^2U(\tilde{x}_{n\gamma}))\,dw_r\right\rangle\right]\\&\le\gamma^3(c_8|x_{n\gamma}-x^*|^2+c_9|\tilde{x}_{n\gamma}-x^*|^2+c_{10}),\end{aligned}$$

where the constants c_8,c_9 *and* c_{10} *are given explicitly in the proof.*

Proof. The proof follows the same lines as in Lemma 3.16 with $\int_{n\gamma}^t M(\tilde{x}_r,\tilde{x}_{n\gamma})\,dw_r$ replaced by $\int_{n\gamma}^t(\nabla^2U(\tilde{x}_r)-\nabla^2U(\tilde{x}_{n\gamma}))\,dw_r$, thus, the main focus here is to provide explicit constants. For any $t\in[n\gamma,(n+1)\gamma)$, one observes that

$$\begin{aligned}&\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left\langle\int_{n\gamma}^t(\nabla U(x_r)-\nabla U(\tilde{x}_r))\,dr,\int_{n\gamma}^t(\nabla^2U(\tilde{x}_r)-\nabla^2U(\tilde{x}_{n\gamma}))\,dw_r\right\rangle\right]\\&=\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left\langle\int_{n\gamma}^t\left\{\nabla U(x_r)-\nabla U(x_{n\gamma})-(\nabla U(\tilde{x}_r)-\nabla U(\tilde{x}_{n\gamma}))\right.\right.\right.\\&\qquad\left.-\sqrt{2}\int_{n\gamma}^r\nabla^2U(x_{n\gamma})\,dw_s+\sqrt{2}\int_{n\gamma}^r\nabla^2U(\tilde{x}_{n\gamma})\,dw_s\right\}dr,\\&\qquad\left.\left.\int_{n\gamma}^t(\nabla^2U(\tilde{x}_r)-\nabla^2U(\tilde{x}_{n\gamma}))\,dw_r\right\rangle\right]\\&\quad+\sqrt{2}\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left\langle\int_{n\gamma}^t\int_{n\gamma}^r(\nabla^2U(x_{n\gamma})-\nabla^2U(\tilde{x}_{n\gamma}))\,dw_s\,dr,\right.\right.\\&\qquad\left.\left.\int_{n\gamma}^t(\nabla^2U(\tilde{x}_r)-\nabla^2U(\tilde{x}_{n\gamma}))\,dw_r\right\rangle\right].\end{aligned}\tag{3.45}$$

By applying Itô's formula to $\nabla^2U(\tilde{x}_r)$, the second term in (3.45) can be estimated as

$$\begin{aligned}&\sqrt{2}\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left\langle\int_{n\gamma}^t\int_{n\gamma}^r(\nabla^2U(x_{n\gamma})-\nabla^2U(\tilde{x}_{n\gamma}))\,dw_s\,dr,\right.\right.\\&\qquad\left.\left.\int_{n\gamma}^t(\nabla^2U(\tilde{x}_r)-\nabla^2U(\tilde{x}_{n\gamma}))\,dw_r\right\rangle\right]\\&=\sqrt{2}\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\sum_{i=1}^d\left(\int_{n\gamma}^t\sum_{l=1}^d\int_{n\gamma}^r(\nabla^2U^{(i,l)}(x_{n\gamma})-\nabla^2U^{(i,l)}(\tilde{x}_{n\gamma}))\,dw_s^{(l)}\,dr\right)\right.\\&\quad\times\left(\sum_{j=1}^d\int_{n\gamma}^t\left(-\int_{n\gamma}^r\sum_{k=1}^d\frac{\partial^3U(\tilde{x}_s)}{\partial x^{(i)}\partial x^{(j)}\partial x^{(k)}}\nabla\tilde{U}^{(k)}(s,\tilde{x}_{n\gamma})\,ds\right.\right.\\&\quad\left.\left.\left.+\int_{n\gamma}^r\sum_{k=1}^d\frac{\partial^4U(\tilde{x}_s)}{\partial x^{(i)}\partial x^{(j)}\partial x^{(k)}\partial x^{(k)}}\,ds+\sqrt{2}\int_{n\gamma}^r\sum_{k=1}^d\frac{\partial^3U(\tilde{x}_s)}{\partial x^{(i)}\partial x^{(j)}\partial x^{(k)}}\,dw_s^{(k)}\right)\,dw_r^{(j)}\right)\right]\\&\le\frac{1}{2}\sum_{i=1}^d\mathbb{E}^{\mathscr{F}_{n\gamma}}\left|\int_{n\gamma}^t\sum_{l=1}^d\int_{n\gamma}^r(\nabla^2U^{(i,l)}(x_{n\gamma})-\nabla^2U^{(i,l)}(\tilde{x}_{n\gamma}))\,dw_s^{(l)}\,dr\right|^2\end{aligned}$$

$$
\begin{aligned}
&+\sum_{i=1}^d \mathbb{E}^{\mathscr{F}_{n\gamma}} \left| \sum_{j=1}^d \int_{n\gamma}^t \left(-\int_{n\gamma}^r \sum_{k=1}^d \frac{\partial^3 U(\tilde{x}_s)}{\partial x^{(i)} \partial x^{(j)} \partial x^{(k)}} \nabla \tilde{U}^{(k)}(s, \tilde{x}_{n\gamma})\, ds \right.\right. \\
&\qquad\qquad \left.\left. + \int_{n\gamma}^r \sum_{k=1}^d \frac{\partial^4 U(\tilde{x}_s)}{\partial x^{(i)} \partial x^{(j)} \partial x^{(k)} \partial x^{(k)}}\, ds \right) dw_r^{(j)} \right|^2 \\
&\le 2L_2^2 |x_{n\gamma} - x^*|^2 d\gamma^3 \\
&\quad + \gamma^3((2L_2^2 + 6L_1^2L_2^2 + 12L_1^4L_2^2)d|\tilde{x}_{n\gamma} - x^*|^2 + 2L^2d^4 + 12L_2^4d^3 + 12L_1^2L_2^2d^2). \qquad (3.46)
\end{aligned}
$$

where the first inequality holds due to Young's inequality and the fact that for any $i, l, j, k = 1, \ldots, d$

$$
\begin{aligned}
\mathbb{E}^{\mathscr{F}_{n\gamma}} \Bigg[&\int_{n\gamma}^t \int_{n\gamma}^r (\nabla^2 U^{(i,l)}(x_{n\gamma}) - \nabla^2 U^{(i,l)}(\tilde{x}_{n\gamma}))\, dw_s^{(l)}\, dr \\
&\int_{n\gamma}^t \int_{n\gamma}^r \sqrt{2} \frac{\partial^3 U(\tilde{x}_s)}{\partial x^{(i)} \partial x^{(j)} \partial x^{(k)}}\, dw_s^{(k)}\, dw_r^{(j)} \Bigg] = 0,
\end{aligned}
$$

while the last inequality holds due to Young's inequality, results in Appendix B.4 and B.5, and Lemma 3.20. By using Cauchy-Schwarz inequality, (3.45) becomes

$$
\begin{aligned}
&\mathbb{E}^{\mathscr{F}_{n\gamma}} \left[\left\langle \int_{n\gamma}^t (\nabla U(x_r) - \nabla U(\tilde{x}_r))\, dr, \int_{n\gamma}^t (\nabla^2 U(\tilde{x}_r) - \nabla^2 U(\tilde{x}_{n\gamma}))\, dw_r \right\rangle \right] \\
&\le \sqrt{\mathbb{E}^{\mathscr{F}_{n\gamma}} \left| \int_{n\gamma}^t (\nabla^2 U(\tilde{x}_r) - \nabla^2 U(\tilde{x}_{n\gamma}))\, dw_r \right|^2} \left(\mathbb{E}^{\mathscr{F}_{n\gamma}} \left[\gamma \int_{n\gamma}^t |\nabla U(x_r) - \nabla U(x_{n\gamma}) \right.\right. \\
&\left.\left. -(\nabla U(\tilde{x}_r) - \nabla U(\tilde{x}_{n\gamma})) - \sqrt{2} \int_{n\gamma}^r \nabla^2 U(x_{n\gamma})\, dw_s + \sqrt{2} \int_{n\gamma}^r \nabla^2 U(\tilde{x}_{n\gamma})\, dw_s|^2\, dr \right] \right)^{1/2} \\
&\quad + 2L_2^2 |x_{n\gamma} - x^*|^2 d\gamma^3 \\
&\quad + \gamma^3((2L_2^2 + 6L_1^2L_2^2 + 12L_1^4L_2^2)d|\tilde{x}_{n\gamma} - x^*|^2 + 2L^2d^4 + 12L_2^4d^3 + 12L_1^2L_2^2d^2).
\end{aligned}
$$

Then, to estimate the first term of (3.45), one applies Itô's formula to $\nabla U(x_r) - \nabla U(x_{n\gamma})$ and $\nabla U(\tilde{x}_r) - \nabla U(\tilde{x}_{n\gamma})$ to obtain, almost surely

$$
\begin{aligned}
&\left(\mathbb{E}^{\mathscr{F}_{n\gamma}} \left[\gamma \int_{n\gamma}^t |\nabla U(x_r) - \nabla U(x_{n\gamma}) - (\nabla U(\tilde{x}_r) - \nabla U(\tilde{x}_{n\gamma})) \right.\right. \\
&\qquad \left.\left. -\sqrt{2} \int_{n\gamma}^r \nabla^2 U(x_{n\gamma})\, dw_s + \sqrt{2} \int_{n\gamma}^r \nabla^2 U(\tilde{x}_{n\gamma})\, dw_s|^2\, dr \right] \right)^{1/2} \\
&\le 2 \left(\mathbb{E}^{\mathscr{F}_{n\gamma}} \left[\gamma^2 \int_{n\gamma}^t \int_{n\gamma}^r \left| \nabla^2 U(x_s) \nabla U(x_s) - \vec{\Delta}(\nabla U)(x_s) \right|^2 ds\, dr \right] \right. \\
&\qquad + \mathbb{E}^{\mathscr{F}_{n\gamma}} \left[\gamma^2 \int_{n\gamma}^t \int_{n\gamma}^r \left| \nabla^2 U(\tilde{x}_s) \nabla \tilde{U}(s, \tilde{x}_{n\gamma}) - \vec{\Delta}(\nabla U)(\tilde{x}_s) \right|^2 ds\, dr \right] \\
&\qquad + 2d\gamma \int_{n\gamma}^t \int_{n\gamma}^r \mathbb{E}^{\mathscr{F}_{n\gamma}} \left[\left| \nabla^2 U(x_s) - \nabla^2 U(x_{n\gamma}) \right|^2 \right] ds\, dr \\
&\qquad \left. +2d\gamma \int_{n\gamma}^t \int_{n\gamma}^r \mathbb{E}^{\mathscr{F}_{n\gamma}} \left[\left| \nabla^2 U(\tilde{x}_s) - \nabla^2 U(\tilde{x}_{n\gamma}) \right|^2 \right] ds\, dr \right)^{1/2} \\
&\le 2 \left(\gamma^2 \int_{n\gamma}^t \int_{n\gamma}^r \mathbb{E}^{\mathscr{F}_{n\gamma}} \left[2L_1^4 |x_s - x^*|^2 + 2L_2^2 d^2 \right] ds\, dr \right. \\
&\qquad + \gamma^2 \int_{n\gamma}^t \int_{n\gamma}^r \mathbb{E}^{\mathscr{F}_{n\gamma}} \left[2L_1^2 \left| \nabla \tilde{U}(s, \tilde{x}_{n\gamma}) \right|^2 + 2L_2^2 d^2 \right] ds\, dr
\end{aligned}
$$

$$
\begin{aligned}
&\quad + 2d\gamma \int_{n\gamma}^{t}\int_{n\gamma}^{r} \mathbb{E}^{\mathscr{F}_{n\gamma}}\left[L_2^2\left|x_s - x_{n\gamma}\right|^2\right] ds\, dr \\
&\quad +2d\gamma \int_{n\gamma}^{t}\int_{n\gamma}^{r} \mathbb{E}^{\mathscr{F}_{n\gamma}}\left[L_2^2\left|\tilde{x}_s - \tilde{x}_{n\gamma}\right|^2\right] ds\, dr\Bigg)^{1/2} \\
&\leq 2\gamma^2\left(2L_1^4|x_{n\gamma} - x^*|^2 + 4\gamma L_1^4 d + 2L_2^2 d^2\right. \\
&\quad + (6L_1^4 + 12\gamma^2 L_1^6)|\tilde{x}_{n\gamma} - x^*|^2 + 12\gamma^2 L_1^2 L_2^2 d^2 + 12\gamma L_1^4 d + 2L_2^2 d^2 \\
&\quad + 2d\left(2\gamma L_1^2 L_2^2 |x_{n\gamma} - x^*|^2 + 4\gamma^2 L_1^2 L_2^2 d + 4L_2^2 d\right) \\
&\quad \left.+2d\left(L_2^2 c_1 |\tilde{x}_{n\gamma} - x^*|^2 + L_2^2 c_2\right)\right)^{1/2},
\end{aligned}
$$

where the first inequality holds due to Cauchy-Schwarz inequality and Young's inequality, the second inequality holds by using (3.33) and **B-5**, while the last inequality is obtained due to Lemma 3.20 and 3.21. Finally by using Young's inequality, one obtains

$$
\begin{aligned}
&\mathbb{E}^{\mathscr{F}_{n\gamma}}\left[\left\langle \int_{n\gamma}^{t} (\nabla U(x_r) - \nabla U(\tilde{x}_r))\, dr, \int_{n\gamma}^{t} (\nabla^2 U(\tilde{x}_r) - \nabla^2 U(\tilde{x}_{n\gamma}))\, dw_r \right\rangle\right] \\
&\leq \gamma^3 (c_8 |x_{n\gamma} - x^*|^2 + c_9 |\tilde{x}_{n\gamma} - x^*|^2 + c_{10}),
\end{aligned}
$$

where $c_8 = 2L_1^4 + 4L_1^2 L_2^2 d + 2L_2^2 d$, $c_9 = (4L_2^2 c_1 + 2L_2^2 + 6L_1^2 L_2^2 + 12L_1^4 L_2^2)d + 6L_1^4 + 12L_1^6$ and $c_{10} = 2L^2 d^4 + 4L_2^2 c_2 d + 12L_2^4 d^3 + 32L_1^2 L_2^2 d^2 + 12L_2^2 d^2 + 16L_1^4 d$. □

Proof of Theorem 3.6. Note that in the Lipschitz case, there are restrictions for the stepsize $\gamma \in \left(0, \frac{1}{\tilde{m}} \wedge \frac{4}{5(m+L_1)} \wedge \frac{\sqrt{m}}{(m+L_1)\sqrt{L_1}} \wedge \sqrt[3]{\frac{1}{(m+L_1)L_1^2}}\right)$. Consider the synchronous coupling of x_t and $\tilde{x}_t$ for $t \geq 0$, where $\tilde{x}_t$ is defined by (3.36). Let $(x_0, \tilde{x}_0)$ distributed according to ζ_0, where $\zeta_0 = \pi \otimes \delta_x$ for all $x \in \mathbb{R}^d$. Define $e_t = x_t - \tilde{x}_t$, for all $t \in [n\gamma, (n+1)\gamma)$, $n \in \mathbb{N}$. By Itô's formula, one obtains, almost surely,

$$
|e_t|^2 = |e_{n\gamma}|^2 - 2\int_{n\gamma}^{t} \left\langle e_s, \nabla U(x_s) - \nabla \tilde{U}(s, \tilde{x}_{n\gamma})\right\rangle ds.
$$

Then, taking the expectation and taking the derivative on both sides yield

$$
\begin{aligned}
\frac{d}{dt}\mathbb{E}\left[|e_t|^2\right] &= -2\mathbb{E}\left[\left\langle e_t, \nabla U(x_t) - \nabla \tilde{U}(t, \tilde{x}_{n\gamma})\right\rangle\right] \\
&= 2\mathbb{E}\left[\langle e_t, -(\nabla U(x_t) - \nabla U(\tilde{x}_t))\rangle\right] \\
&\quad + 2\mathbb{E}\left[\langle e_t, -(\nabla U(\tilde{x}_t) - \nabla U(\tilde{x}_{n\gamma}) - \nabla U_1(t, \tilde{x}_{n\gamma}) - \nabla U_2(t, \tilde{x}_{n\gamma}))\rangle\right].
\end{aligned}
$$

By applying Itô's formula to $\nabla U(\tilde{x}_t) - \nabla U(\tilde{x}_{n\gamma})$, and by calculating $\nabla U(\tilde{x}_t) - \nabla U(\tilde{x}_{n\gamma}) - \nabla U_1(t, \tilde{x}_{n\gamma}) - \nabla U_2(t, \tilde{x}_{n\gamma})$, one obtains (3.43). Substituting (3.43) into the above equation and by using **B-3** yield

$$
\begin{aligned}
\frac{d}{dt}\mathbb{E}\left[|e_t|^2\right] &\leq -2m\mathbb{E}\left[|e_t|^2\right] \\
&\quad + 2\mathbb{E}\left[|e_t|\left|\int_{n\gamma}^{t} (\nabla^2 U(\tilde{x}_r) - \nabla^2 U(\tilde{x}_{n\gamma}))\nabla U(\tilde{x}_{n\gamma})\, dr\right|\right] \\
&\quad + 2\mathbb{E}\left[|e_t|\left|\int_{n\gamma}^{t} \nabla^2 U(\tilde{x}_r)(\nabla U_1(r, \tilde{x}_{n\gamma}) + \nabla U_2(r, \tilde{x}_{n\gamma}))\, dr\right|\right] \\
&\quad + 2\sqrt{2}\mathbb{E}\left[\left\langle e_t, \left(-\int_{n\gamma}^{t} (\nabla^2 U(\tilde{x}_r) - \nabla^2 U(\tilde{x}_{n\gamma}))\, dw_r\right)\right\rangle\right] \\
&\quad + 2\mathbb{E}\left[|e_t|\left|\int_{n\gamma}^{t} (\vec{\Delta}(\nabla U)(\tilde{x}_r) - \vec{\Delta}(\nabla U)(\tilde{x}_{n\gamma}))\, dr\right|\right] \qquad (3.47) \\
&\leq -(2m - 3\varepsilon)\mathbb{E}\left[|e_t|^2\right]
\end{aligned}
$$

$$
\begin{aligned}
&+ \frac{1}{\varepsilon}\mathbb{E}\left[\left|\int_{n\gamma}^{t} (\nabla^2 U(\tilde{x}_r) - \nabla^2 U(\tilde{x}_{n\gamma}))\nabla U_\gamma(\tilde{x}_{n\gamma})\, dr\right|^2\right] \\
&+ \frac{1}{\varepsilon}\mathbb{E}\left[\left|\int_{n\gamma}^{t} \nabla^2 U(\tilde{x}_r)(\nabla U_{1,\gamma}(r, \tilde{x}_{n\gamma}) + \nabla U_{2,\gamma}(r, \tilde{x}_{n\gamma}))\, dr\right|^2\right] \\
&+ \frac{1}{\varepsilon}\mathbb{E}\left[\left|\int_{n\gamma}^{t} (\vec{\Delta}(\nabla U)(\tilde{x}_r) - \vec{\Delta}(\nabla U)(\tilde{x}_{n\gamma}))\, dr\right|^2\right] \\
&- 2\sqrt{2}\mathbb{E}\left[\left\langle e_t - e_{n\gamma}, \int_{n\gamma}^{t} (\nabla^2 U(\tilde{x}_r) - \nabla^2 U(\tilde{x}_{n\gamma}))\, dw_r \right\rangle\right] \\
&- 2\sqrt{2}\mathbb{E}\left[\left\langle e_{n\gamma}, \int_{n\gamma}^{t} (\nabla^2 U(\tilde{x}_r) - \nabla^2 U(\tilde{x}_{n\gamma}))\, dw_r \right\rangle\right],
\end{aligned}
$$

where the second inequality holds due to Young's inequality and the last term is zero. Then, by using the results in Lemma 3.21 and by taking $\varepsilon = \frac{m}{4}$, one obtains

$$
\begin{aligned}
&\frac{d}{dt}\mathbb{E}\left[|e_t|^2\right] \\
&\leq -m\mathbb{E}\left[|e_t|^2\right] \\
&\quad + \frac{4}{m}\gamma^3\mathbb{E}\left[(c_1 L_1^2 L_2^2 |\tilde{x}_{n\gamma} - x^*|^4 + c_2 L_1^2 L_2^2 |\tilde{x}_{n\gamma} - x^*|^2)\right] \\
&\quad + \frac{4}{m}\gamma^3\mathbb{E}\left[(4\gamma L_1^6 |\tilde{x}_{n\gamma} - x^*|^2 + 4\gamma L_1^2 L_2^2 d^2 + 4dL_1^4)\right] \\
&\quad + \frac{4}{m}\gamma^3\mathbb{E}\left[(d^{3/2} L c_1 |\tilde{x}_{n\gamma} - x^*|^2 + d^{3/2} L c_2)\right] \\
&\quad + 2\sqrt{2}\mathbb{E}\left[\left\langle \int_{n\gamma}^{t} (\nabla U(x_r) - \nabla U(\tilde{x}_r))\, dr, \int_{n\gamma}^{t} (\nabla^2 U(\tilde{x}_r) - \nabla^2 U(\tilde{x}_{n\gamma}))\, dw_r \right\rangle\right] \\
&\quad + 2\sqrt{2}\mathbb{E}\left[\left\langle \int_{n\gamma}^{t} (\nabla U(\tilde{x}_r) - \nabla U(\tilde{x}_{n\gamma}) - \nabla U_1(r, \tilde{x}_{n\gamma}) - \nabla U_2(r, \tilde{x}_{n\gamma}))\, dr,\right.\right. \\
&\qquad\qquad\qquad \left.\left. \int_{n\gamma}^{t} (\nabla^2 U(\tilde{x}_r) - \nabla^2 U(\tilde{x}_{n\gamma}))\, dw_r \right\rangle\right] \\
&\leq -m\mathbb{E}\left[|e_t|^2\right] \\
&\quad + \frac{4}{m}\gamma^3\mathbb{E}\left[(c_1 L_1^2 L_2^2 |\tilde{x}_{n\gamma} - x^*|^4 + c_2 L_1^2 L_2^2 |\tilde{x}_{n\gamma} - x^*|^2)\right] \\
&\quad + \frac{4}{m}\gamma^3\mathbb{E}\left[(4\gamma L_1^6 |\tilde{x}_{n\gamma} - x^*|^2 + 4\gamma L_1^2 L_2^2 d + 4dL_1^4)\right] \\
&\quad + \frac{4}{m}\gamma^3\mathbb{E}\left[(d^{3/2} L c_1 |\tilde{x}_{n\gamma} - x^*|^2 + d^{3/2} L c_2)\right] \\
&\quad + 2\sqrt{2}\gamma^3\mathbb{E}\left[(c_8 |x_{n\gamma} - x^*|^2 + c_9 |\tilde{x}_{n\gamma} - x^*|^2 + c_{10})\right] \\
&\quad + 2\gamma^3\mathbb{E}\left[(c_5 |\tilde{x}_{n\gamma} - x^*|^4 + c_6 |\tilde{x}_{n\gamma} - x^*|^2 + c_7)\right] \\
&\quad + 2\gamma^3\mathbb{E}\left[d(L_2^2 c_1 |\tilde{x}_{n\gamma} - x^*|^2 + L_2^2 c_2)\right]
\end{aligned}
$$

where the last inequality holds by using Cauchy-Schwarz inequality, Young's inequality and Lemma 3.21, 3.23, 3.22. Then, after simplification, one obtains

$$
\begin{aligned}
\frac{d}{dt}\mathbb{E}\left[|e_t|^2\right] &\leq -m\mathbb{E}\left[|e_t|^2\right] \\
&\quad + \gamma^3\mathbb{E}\left[(c_{11}|\tilde{x}_{n\gamma} - x^*|^4 + c_{12}|\tilde{x}_{n\gamma} - x^*|^2 + c_{13}|x_{n\gamma} - x^*|^2 + c_{14})\right],
\end{aligned}
$$

where $c_{11} = \frac{4}{m}L_1^2 L_2^2 c_1 + 2c_5$, $c_{12} = \frac{4}{m}(L_1^2 L_2^2 c_2 + 4L_1^6 + d^{3/2} L c_1) + 2\sqrt{2}c_9 + 2c_6 + 2dL_2^2 c_1$, $c_{13} = 2\sqrt{2}c_8$ and $c_{14} = \frac{4}{m}(4L_1^2 L_2^2 d + 4L_1^4 d + d^{3/2} L c_2) + 2\sqrt{2}c_{10} + 2c_7 + 2L_2^2 d c_2$. Then, the application of Gronwall's lemma yields

$$\mathbb{E}\left[|e_t|^2\right] \leq e^{-m(t-n\gamma)}\mathbb{E}\left[|e_{n\gamma}|^2\right] \\ + \gamma^4\mathbb{E}\left[(c_{11}|\tilde{x}_{n\gamma}-x^*|^4 + c_{12}|\tilde{x}_{n\gamma}-x^*|^2 + c_{13}|x_{n\gamma}-x^*|^2 + c_{14})\right].$$

Finally, by induction, Proposition 3.18 and 3.19, one obtains

$$\begin{aligned}
&\mathbb{E}\left[|e_{(n+1)\gamma}|^2\right] \\
&\leq e^{-m\gamma(n+1)}\mathbb{E}\left[|e_0|^2\right] + \frac{\gamma^4 c_{14}}{1-e^{-m\gamma}} + \gamma^4 c_{11}\sum_{k=0}^{n}\mathbb{E}\left[|\tilde{x}_{k\gamma}-x^*|^4\right]e^{-m\gamma(n-k)} \\
&\quad + \gamma^4 c_{12}\sum_{k=0}^{n}\mathbb{E}\left[|\tilde{x}_{k\gamma}-x^*|^2\right]e^{-m\gamma(n-k)} + \gamma^4 c_{13}\sum_{k=0}^{n}\mathbb{E}\left[|x_{k\gamma}-x^*|^2\right]e^{-m\gamma(n-k)} \\
&\leq e^{-m\gamma(n+1)}\mathbb{E}\left[|x_0-\tilde{x}_0|^2\right] + \frac{\gamma^3 e^m}{m}\left(c_{14} + c_{11}\left(\mathbb{E}\left[|\tilde{x}_0-x^*|^4\right] + \frac{8q_2}{\tilde{m}}\right)\right. \\
&\quad \left. + c_{12}\left(\mathbb{E}\left[|\tilde{x}_0-x^*|^2\right] + \frac{q_1}{\tilde{m}}\right) + c_{13}\left(\mathbb{E}\left[|x_0-x^*|^2\right] + 2d\right)\right)
\end{aligned}$$

where the last inequality holds by using $1-e^{-m\gamma} \geq m\gamma e^{-m\gamma}$. The application of Theorem 1 in [15] with the initial distribution ζ_0 yields

$$W_2^2(\delta_x \tilde{R}_\gamma^n, \pi) \leq e^{-mn\gamma}\left(2|x-x^*|^2 + \frac{2d}{m}\right) + \bar{C}\gamma^3,$$

where $\bar{C} = O(d^4)$ with

$$\bar{C} = \frac{e^m}{m}\left(c_{14} + c_{11}\left(|x-x^*|^4 + \frac{8q_2}{\tilde{m}}\right) + c_{12}\left(|x-x^*|^2 + \frac{q_1}{\tilde{m}}\right) + c_{13}\left(\frac{d}{m} + 2d\right)\right)$$

Proof of Corollary 3.7. In the case that the target distribution π is a multivariate Gaussian distribution, by using the same arguments, one notices that for

$$\gamma \in \left(0, \frac{1}{\tilde{m}} \wedge \frac{4}{5(m+L_1)} \wedge \frac{\sqrt{m}}{(m+L_1)\sqrt{L_1}} \wedge \sqrt[3]{\frac{1}{(m+L_1)L_1^2}}\right),$$

Proposition 3.18 holds with $q_1 = (4L_1^2+4)d$. Then, one obtains the following bound

$$\mathbb{E}\left[|e_t|^2\right] \leq e^{-m(t-n\gamma)}\mathbb{E}\left[|e_{n\gamma}|^2\right] + \gamma^4\mathbb{E}\left[\frac{4}{m}\left(4L_1^6|\tilde{x}_{n\gamma}-x^*|^2 + 4dL_1^4\right)\right],$$

which indicates

$$W_2^2(\delta_x \tilde{R}_\gamma^n, \pi) \leq e^{-mn\gamma}\left(2|x-x^*|^2 + \frac{2d}{m}\right) + \tilde{C}\gamma^3,$$

where $\tilde{C} = \frac{16L_1^4 e^m}{m^2}\left(d + L_1^2\left(|x-x^*|^2 + \frac{q_1}{\tilde{m}}\right)\right)$.

3.5.3 Example: Logistic regression with Gaussian prior

We provide an example of the logistic regression in dimension d. Denote by θ_k, $k \in \mathbb{N}$ the k-th iteration of the algorithm (3.8). One observes a sequence of i.i.d. sample $\{(x_i, y_i)\}_{i=1,\ldots,n}$, where $x_i \in \mathbb{R}^d$ and $y_i \in \{0,1\}$ for all i. The likelihood function is given by $p(y_i|x_i,\theta) = (1/(1+e^{-x_i^\top\theta}))^{y_i}(1-1/(1+e^{-x_i^\top\theta}))^{1-y_i}$. Consider a Gaussian prior with mean zero and covariance matrix proportional to the inverse of the matrix $\Sigma_X = \frac{1}{n}\sum_{i=1}^n x_i x_i^\top$. For $\theta \in \mathbb{R}^d$, the gradient $\nabla U(\theta)$ and Hessian $\nabla^2 U(\theta)$ with n data points are

$$\nabla U(\theta) = c\Sigma_X\theta + \sum_{i=1}^n\left(\frac{x_i}{1+e^{-x_i^\top\theta}} - y_i x_i\right), \quad \nabla^2 U(\theta) = c\Sigma_X + \sum_{i=1}^n \frac{x_i x_i^\top e^{-x_i^\top\theta}}{(1+e^{-x_i^\top\theta})^2},$$

where $c > 0$. This implies that $L_1 \le (c+n)\max_i |x_i x_i^\intercal|$ with $|x_i x_i^\intercal|$ the spectral norm of the matrix $x_i x_i^\intercal$ for each i. One notices that $\max_i |x_i x_i^\intercal|$ is much smaller than $\max_i |x_i|^2 = O(d)$ due to the fact that the matrix $x_i x_i^\intercal$ is typically sparse in statistical and machine learning applications. One may refer to dimension reduction techniques in sparse matrices in data science for more discussions, see e.g. [23] and [33].

To calculate the Lipschitz constant L_2 in **B-5**, one denotes $g(\lambda) = \nabla^2 U(\lambda y + (1-\lambda)x)$, for any $x, y \in \mathbb{R}^d$ and $\lambda \in [0,1]$. By fundamental theorem of calculus, one obtains, for any $l = 1, \ldots, d$

$$\begin{aligned} g^{(l,\cdot)}(1) - g^{(l,\cdot)}(0) &= \nabla^2 U^{(l,\cdot)}(y) - \nabla^2 U^{(l,\cdot)}(x) \\ &= \int_0^1 \nabla^2 (\nabla U)^{(l)}(\lambda y + (1-\lambda)x)(y-x)\, d\lambda, \end{aligned}$$

where $\nabla^2(\nabla U)^{(l)}$ is a matrix with (j,k)-th entry $\frac{\partial^3 U}{\partial x^l \partial x^j \partial x^k}$ and for any $\theta \in \mathbb{R}^d$

$$\nabla^2(\nabla U)^{(l)}(\theta) = \sum_{i=1}^n \left(\frac{2x_i^{(l)} x_i x_i^\intercal e^{-2x_i^\intercal \theta}}{(1+e^{-x_i^\intercal \theta})^3} - \frac{x_i^{(l)} x_i x_i^\intercal e^{-x_i^\intercal \theta}}{(1+e^{-x_i^\intercal \theta})^2} \right)$$

Moreover, one notices

$$\begin{aligned} |\nabla^2 U(y) - \nabla^2 U(x)| &\le \left(\sum_{l=1}^d \left| \nabla^2 U^{(l,\cdot)}(y) - \nabla^2 U^{(l,\cdot)}(x) \right|^2 \right)^{1/2} \\ &\le 3n \max_i |x_i||x_i x_i^\intercal||y-x| \end{aligned}$$

which implies $L_2 = 3n \max_i |x_i||x_i x_i^\intercal|$.

Finally, for the constant L in **B-6**, define for any $k = 1, \ldots, d$, $f_k(\lambda) = \nabla^2(\nabla U)^{(k)}(\lambda y + (1-\lambda)x)$, for any $x, y \in \mathbb{R}^d$ and $\lambda \in [0,1]$, and one uses the same technique to obtain, for any $l = 1, \ldots, d$

$$\begin{aligned} f_k^{(l,\cdot)}(1) - f_k^{(l,\cdot)}(0) &= (\nabla^2(\nabla U)^{(k)})^{(l,\cdot)}(y) - (\nabla^2(\nabla U)^{(k)})^{(l,\cdot)}(x) \\ &= \int_0^1 \nabla^2(\nabla^2 U)^{(k,l)}(\lambda y + (1-\lambda)x)(y-x)\, d\lambda. \end{aligned}$$

where $\nabla^2(\nabla^2 U)^{(k,l)}$ is a matrix with (j,m)-th entry $\frac{\partial^4 U}{\partial x^k \partial x^l \partial x^j \partial x^m}$ and for any $\theta \in \mathbb{R}^d$

$$\begin{aligned} \nabla^2(\nabla^2 U)^{(k,l)}(\theta) \le \sum_{i=1}^n \Bigg| & \frac{x_i^{(k)} x_i^{(l)} x_i x_i^\intercal e^{-x_i^\intercal \theta}}{(1+e^{-x_i^\intercal \theta})^2} \\ & - \frac{6x_i^{(k)} x_i^{(l)} x_i x_i^\intercal e^{-2x_i^\intercal \theta}}{(1+e^{-x_i^\intercal \theta})^3} + \frac{6x_i^{(k)} x_i^{(l)} x_i x_i^\intercal e^{-3x_i^\intercal \theta}}{(1+e^{-x_i^\intercal \theta})^4} \Bigg| \end{aligned}$$

Then, one obtains for $k = 1, \ldots, d$,

$$\begin{aligned} & |\nabla^2(\nabla U)^{(k)}(y) - \nabla^2(\nabla U)^{(k)}(x)| \\ &= \left(\sum_{l=1}^d \left| (\nabla^2(\nabla U)^{(k)})^{(l,\cdot)}(y) - (\nabla^2(\nabla U)^{(k)})^{(l,\cdot)}(x) \right|^2 \right)^{1/2} \\ &\le 13n \max_i |x_i^{(k)}||x_i||x_i x_i^\intercal||y-x|, \end{aligned}$$

which implies $L \le 13n \max_i |x_i^{(k)}||x_i||x_i x_i^\intercal|$, and effectively, it has the same dimension dependence as L_2.

Chapter 4

Non-asymptotic estimates for Stochastic Gradient Langevin Dynamics with discontinuity

4.1 Introduction

We consider the optimization problem:

$$\text{minimize} \quad \hat{U}(\theta) := \mathbb{E}[f(\theta, Z)],$$

where $\theta \in \mathbb{R}^d$ and Z is a random element. We aim to generate $\hat{\theta}$ such that the expected excess risk

$$\mathbb{E}[\hat{U}(\hat{\theta})] - \inf_{\theta \in \mathbb{R}^d} \hat{U}(\theta)$$

is minimized. To achieve this, as discussed in Chapter 1, we consider the SGLD algorithm given by

$$\theta_{n+1}^{\gamma} := \theta_n^{\gamma} - \gamma H(\theta_n^{\gamma}, Z_{n+1}) + \sqrt{2\beta^{-1}\gamma}\xi_{n+1}, \quad \theta_0^{\gamma} := \theta_0, \tag{4.1}$$

where θ_0 is an $\mathbb{R}^d$-valued random variable, $\gamma > 0$ is the stepsize, $\beta > 0$, $H : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ is a measurable function satisfying $\nabla \hat{U}(\theta) = \mathbb{E}[H(\theta, Z_0)]$, $(Z_n)_{n\in\mathbb{N}}$ is a sequence of i.i.d. random variables, and $(\xi_n)_{n\in\mathbb{N}}$ is a sequence of standard independent d-dimensional Gaussian random variables. Moreover, the SGLD algorithm (4.1) can be viewed as the discretization of the the Langevin SDE given by

$$d\hat{Y}_t = -h(\hat{Y}_t)dt + \sqrt{2\beta^{-1}}dw_t, \quad \hat{Y}_0 = \theta_0, \tag{4.2}$$

where $h := \nabla \hat{U}$ and $(w_t)_{t\geq 0}$ represents the standard Brownian motion. Under mild conditions, it can be shown that the SDE (4.2) admits an unique invariant measure $\pi_\beta(\theta) \propto \exp(-\beta \hat{U}(\theta))$ with $\beta > 0$.

In this chapter, we aim to establish non-asymptotic error bounds for the SGLD algorithm (4.1) with discontinuous gradient $H(\theta, x)$. More precisely, H is decomposed into two parts F and G, where $F : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ is locally Lipschitz continuous in x and $G : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ is bounded in θ. Furthermore, H is assumed to satisfy a conditional Lipschitz-continuity (CLC) property proposed in [8], which is given explicitly in assumption **C-3** below. By using similar techniques as in [9] and [8], non-asymptotic results in Wasserstein-1 and Wasserstein-2 distance between the law of the SGLD algorithm (4.1) and the target distribution π_β are obtained in both convex and non-convex case. To illustrate the applicability of the SGLD algorithm with discontinuous H, examples from quantile estimation, (modified) Kohonen algorithm and VaR-CVaR algorithm are presented. Numerical experiments are implemented and the results support our theoretical findings.

The chapter is based on my joint work [48], and it is organised as follows. Section 4.2

presents the assumptions and main results. In Section 4.3, the proofs for the main theorem in the non-convex case are provided, which is followed by the proofs for the results in the convex case in Section 4.4. Practical examples are presented in Section 4.5 while auxiliary results are provided in Appendix C.

4.2 Main results

Denote by $(\mathscr{G}_n)_{n\in\mathbb{N}}$ a given filtration representing the flow of past information. $(Z_n)_{n\in\mathbb{N}}$ is an $\mathbb{R}^m$-valued, $(\mathscr{G}_n)$-adapted process. It is assumed throughout the chapter that θ_0, $\mathscr{G}_\infty$ and $(\xi_n)_{n\in\mathbb{N}}$ are independent. Moreover, the following assumptions are considered:

C-1 Let $H : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ take the form

$$H(\theta, z) = F(\theta, z) + G(\theta, z), \quad \theta \in \mathbb{R}^d, \quad z \in \mathbb{R}^m,$$

where $F : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ and $G : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ satisfy the following:

(i) $F : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ is jointly Lipschitz continuous in both variables, i.e. there exist $L_1, L_2 > 0$, $\rho \geq 0$ such that for any $\theta, \theta' \in \mathbb{R}^d$, $z, z' \in \mathbb{R}^m$,

$$|F(\theta, z) - F(\theta', z')| \leq (1 + |z| + |z'|)^\rho (L_1|\theta - \theta'| + L_2|z - z'|).$$

(ii) $G(\theta, z) : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ is bounded in θ, i.e. there exist $K_1 : \mathbb{R}^m \to \mathbb{R}_+$ such that for any $\theta \in \mathbb{R}^d$, $z \in \mathbb{R}^m$,

$$|G(\theta, z)| \leq K_1(z).$$

C-2 We assume the inital value θ_0 satisfies $|\theta_0| \in \mathscr{L}^4$. The process $(Z_n)_{n\in\mathbb{N}}$ is i.i.d. with $|Z_0| \in \mathscr{L}^{4\rho+4}$ and $|K_1(Z_0)| \in \mathscr{L}^4$. Moreover, it satisfies

$$\mathbb{E}[H(\theta, Z_0)] = h(\theta)$$

Remark 4.1. *By **C-1**, for all $\theta \in \mathbb{R}^d$ and $z \in \mathbb{R}^m$,*

$$|H(\theta, z)| \leq (1 + |z|)^{\rho+1}(L_1|\theta| + L_2) + F_*(z),$$

where $F_*(z) = |F(0, 0)| + K_1(z)$. For any $z \in \mathbb{R}^m$, $\rho \geq 0$, denote by

$$K_\rho(z) = (1 + 2|z|)^{4\rho+4}. \tag{4.3}$$

*One notices that by **C-2**, $\mathbb{E}[K_\rho(Z_0)]$ is well defined.*

C-3 There exists a positive constant $L > 0$ such that, for all $\theta, \theta' \in \mathbb{R}^d$,

$$\mathbb{E}[|H(\theta, Z_0) - H(\theta', Z_0)|] \leq L|\theta - \theta'|.$$

Remark 4.2. *By **C-3**, one obtains, for all $\theta, \theta' \in \mathbb{R}^d$,*

$$|h(\theta) - h(\theta')| \leq L|\theta - \theta'|. \tag{4.4}$$

Remark 4.3. *One notes that **C-3** is satisfied for a wide class of $(Z_n)_{n\in\mathbb{N}}$, see Section 4.5 for the examples. Here, for the illustrative purpose, we consider the following. Suppose $G(\theta, z) = \sum_{j=1}^N \dot{g}_j(\theta, z)\mathbb{1}_{\bigcap_{i=1}^m \{z^{(i)} \in I_{i,j}(\theta)\}}$ is a lower semi-continuous function, where $N \in \mathbb{N}^*$, $\dot{g}_j : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ are bounded and jointly Lipschitz continuous functions, i.e. there exist $L_3, L_4, K_2 > 0$ such that for any $\theta, \theta' \in \mathbb{R}^d$, $z, z' \in \mathbb{R}^m$, $j = 1, \ldots, N$*

$$|\dot{g}_j(\theta, z) - \dot{g}_j(\theta', z')| \leq (1 + |z| + |z'|)^\rho (L_3|\theta - \theta'| + L_4|z - z'|), \quad |\dot{g}_j(\theta, z)| \leq K_2.$$

Moreover, the intervals $I_{i,j}(\theta)$ take the form $(-\infty, \bar{g}_j^{(i)}(\theta))$, $(\bar{g}_j^{(i)}(\theta), \infty)$ or $(\tilde{g}_j^{(i)}(\theta), \hat{g}_j^{(i)}(\theta))$, and $\bar{g}_j^{(i)}, \tilde{g}_j^{(i)}, \hat{g}_j^{(i)} : \mathbb{R}^d \to \mathbb{R}$ are Lipschitz continuous functions. In this case, it is enough to require *the marginal density function $f_{Z_0^{(i)}}$ of $Z_0^{(i)}$, for any $i = 1, \ldots, m$, are continuous and bounded. Then, the property stated in **C-3** holds.*

Proof. See Appendix C.1. □

4.2.1 Nonconvex case

Further to the assumptions above, we consider the following dissipativity condition on F, which can be viewed as a generalization of the convexity assumption.

C-4 There exist $A : \mathbb{R}^m \to \mathbb{R}^{d \times d}$, $b : \mathbb{R}^m \to \mathbb{R}$ such that for any $z \in \mathbb{R}^m$, $y \in \mathbb{R}^d$,

$$\langle y, A(z)y \rangle \geq 0$$

and for all $\theta \in \mathbb{R}^d$ and $z \in \mathbb{R}^m$,

$$\langle F(\theta, z), \theta \rangle \geq \langle \theta, A(z)\theta \rangle - b(z).$$

The smallest eigenvalue of $\mathbb{E}[A(Z_0)]$ is a positive real number $a > 0$ and $E[b(Z_0)] = b > 0$.

Remark 4.4. *Compared to [42, **(A.3)**], **C-4** is a relaxed (local) dissipativity condition on F in the sense that it allows the non-uniform dependence in z. See [53] for more discussions on the local conditions.*

Define first

$$\gamma_{\max} = \min \left\{ \frac{\min\{a, a^{1/3}\}}{24(1 + L_1)^2 \mathbb{E}\left[K_\rho(Z_0)\right]}, \frac{1}{4a} \right\}, \tag{4.5}$$

where L_1, a are given in **C-1** and **C-4** respectively, and $K_\rho(z)$ for any $z \in \mathbb{R}^m$ is defined in (4.3).

Remark 4.5. *Throughout this chapter, the constant $C_* > 0$ may take different values at different places, but it is always independent of $n \in \mathbb{N}$, $\beta > 0$ and $d \geq 1$.*

Theorem 4.6. *Assume **C-1**, **C-2**, **C-3** and **C-4** hold. Then, for any $n \in \mathbb{N}$, $0 < \gamma \leq \gamma_{\max}$, there exist constants $C_0, C_1, C_2 > 0$ such that,*

$$W_1(\mathcal{L}(\theta_n^\gamma), \pi_\beta) \leq C_1 e^{-C_0 \gamma n}(\mathbb{E}[|\theta_0|^4] + 1) + C_2\sqrt{\gamma}, \ n \in \mathbb{N}, \tag{4.6}$$

where $C_0 = \dot{c}/2$, $C_1 = O\left(e^{C_* d^2/\beta^2}\left(1 + \frac{1}{1-e^{-\dot{c}}}\right)\right)$ and $C_2 = O\left(e^{C_* d^2/\beta^2}\left(1 + \frac{1}{1-e^{-\dot{c}}}\right)\right)$. The explicit expressions of the constants are provided in (4.27).

Remark 4.7. *The constant $\dot{c}$ is the contraction rate for the Langevin SDE (4.2) in $\tilde{W}_{1,2}$, see Proposition 4.23, which is obtained by using [17, Theorem 2.2]. The definition of $\tilde{W}_{1,2}$ is given in (4.18). Moreover, one notes that C_1 and C_2 are $O\left(e^{C_* d^2/\beta^2}\left(1 + \frac{1}{1-e^{-\dot{c}}}\right)\right)$, and this is due to the fact that the worst case scenario is considered in the analysis of [17, Theorem 2.2]. However, the exponential dependence of C_1, C_2 on d can be lessened by tuning β.*

Theorem 4.6 provides the rate of convergence between the law of the SGLD algorithm (4.1) and the target distribution π_β in Wasserstein-1 distance. Furthermore, a convergence result in Wasserstein-2 distance with reduced rate can be obtained.

Corollary 4.8. *Assume **C-1**, **C-2**, **C-3** and **C-4** hold. Then, for any $n \in \mathbb{N}$, $0 < \gamma \leq \gamma_{\max}$ given in (4.5), there exist constants $C_3, C_4, C_5 > 0$ such that,*

$$W_2(\mathcal{L}(\theta_n^\gamma), \pi_\beta) \leq C_4 e^{-C_3 \gamma n}(\mathbb{E}[|\theta_0|^4] + 1) + C_5 \gamma^{1/4}, \ n \in \mathbb{N},$$

where $C_3 = \dot{c}/4$, $C_4 = O\left(e^{C_* d^2/\beta^2}\left(1 + \frac{1}{1-e^{-\dot{c}/2}}\right)\right)$ and $C_5 = O\left(e^{C_* d^2/\beta^2}\left(1 + \frac{1}{1-e^{-\dot{c}/2}}\right)\right)$ with the explicit expressions given in (4.28).

By using the convergence result in Wasserstein-2 distance as presented in Corollary 4.8, one can obtain an upper bound for the expected excess risk $\mathbb{E}[\hat{U}(\hat{\theta})] - \inf_{\theta\in\mathbb{R}^d} \hat{U}(\theta)$.

Corollary 4.9. *Assume **C-1**, **C-2**, **C-3** and **C-4** hold. Then, for any $n \in \mathbb{N}$, $0 < \gamma \leq \gamma_{\max}$ given in* (4.5)*, there exist constants $\hat{C}_0, \hat{C}_1, \hat{C}_2, \hat{C}_3 > 0$ such that the expected excess risk can be estimated as*

$$\mathbb{E}[\hat{U}(\hat{\theta})] - \inf_{\theta\in\mathbb{R}^d} \hat{U}(\theta) \leq \hat{C}_1 e^{-\hat{C}_0\gamma n} + \hat{C}_2\gamma^{1/4} + \hat{C}_3/\beta,$$

where $\hat{\theta} = \theta^\gamma_n$, and $\hat{C}_0 = \dot{c}/4$, $\hat{C}_1 = O\left(e^{C_* d^2/\beta^2}\left(1 + \frac{1}{1-e^{-\dot{c}/2}}\right)\right)$, $\hat{C}_2 = O\left(e^{C_* d^2/\beta^2}\left(1 + \frac{1}{1-e^{-\dot{c}/2}}\right)\right)$, $\hat{C}_3 = O\left(d\log\left(C_*\left(\frac{\beta}{d}+1\right)\right)\right)$. The explicit expressions of the constants are provided in (4.30) *and* (4.31)*.*

4.2.2 Convex case

Recall **C-1**, where it is assumed $H = F + G$. In this section, we present (improved) convergence results of the SGLD algorithm (4.1) under the convexity condition of F and G.

In the case that F satisfies a convexity condition but not G, the result in Theorem 4.6 can be recovered.

C-5 There exist $\hat{A}_1 : \mathbb{R}^m \to \mathbb{R}^{d\times d}$ such that for any $z \in \mathbb{R}^m$, $y \in \mathbb{R}^d$,

$$\langle y, \hat{A}_1(z)y\rangle \geq 0$$

and for each $\theta, \theta' \in \mathbb{R}^d$, $z \in \mathbb{R}^m$,

$$\langle F(\theta, z) - F(\theta', z), \theta - \theta'\rangle \geq \langle \theta - \theta', \hat{A}_1(z)(\theta - \theta')\rangle.$$

The smallest eigenvalue of $\mathbb{E}[\hat{A}_1(Z_0)]$ is a positive real number $\hat{a}_1 > \epsilon$ with $\epsilon > 0$.

Remark 4.10. *By **C-1** and **C-5**, one obtains, for any $\theta \in \mathbb{R}^d$ and $z \in \mathbb{R}^m$,*

$$\langle F(\theta, z), \theta\rangle \geq \langle \theta, \hat{A}^*_1(z)\theta\rangle - \hat{b}(z),$$

where $\hat{A}^*_1(z) = \hat{A}_1(z) - \epsilon\mathbf{I}_d$ and $\hat{b}(z) = (L_2(1 + |z|)^{\rho+1} + |F(0, 0)|)^2/(4\epsilon)$.

Proof. See Appendix C.2. ☐

Corollary 4.11. *Assume **C-1**, **C-2**, **C-3** and **C-5** hold. Then, for any $n \in \mathbb{N}$, $0 < \gamma \leq \gamma^*_{\max}$, where*

$$\gamma^*_{\max} = \min\left\{\frac{\min\{a^*, (a^*)^{1/3}\}}{24(1 + L_1)^2\mathbb{E}\left[K_\rho(Z_0)\right]}, \frac{1}{4a^*}\right\}$$

with $a^* = \hat{a}_1 - \epsilon$, there exist constants $C^*_0, C^*_1, C^*_2 > 0$ such that,

$$W_1(\mathcal{L}(\theta^\gamma_n), \pi_\beta) \leq C^*_1 e^{-C^*_0\gamma n}(\mathbb{E}[|\theta_0|^4] + 1) + C^*_2\sqrt{\gamma},\ n \in \mathbb{N}. \tag{4.7}$$

If G is assumed to be convex in addition to **C-5**, then it can be shown that the rate of convergence is 1/2 in Wasserstein-2 distance between the law of the SGLD algorithm (4.1) and the target distribution π_β, which appeared to be optimal, see [3, Example 3.4].

C-6 There exist $\hat{A}_2 : \mathbb{R}^m \to \mathbb{R}^{d\times d}$ such that for any $z \in \mathbb{R}^m$, $y \in \mathbb{R}^d$,

$$\langle y, \hat{A}_2(z)y\rangle \geq 0$$

and for each $\theta, \theta' \in \mathbb{R}^d$, $z \in \mathbb{R}^m$,

$$\langle G(\theta, z) - G(\theta', z), \theta - \theta'\rangle \geq \langle \theta - \theta', \hat{A}_2(z)(\theta - \theta')\rangle.$$

The smallest eigenvalue of $\mathbb{E}[\hat{A}_2(Z_0)]$ is a positive real number $\hat{a}_2 > 0$.

Remark 4.12. *By **C-5** and **C-6**, one obtains, for each* $\theta, \theta' \in \mathbb{R}^d$, $z \in \mathbb{R}^m$,

$$\langle H(\theta, z) - H(\theta', z), \theta - \theta' \rangle \geq \langle \theta - \theta', \hat{A}(z)(\theta - \theta') \rangle,$$

where $\hat{A}(z) = \hat{A}_1(z) + \hat{A}_2(z)$*. One notes that the condition above is similar to Assumption 3.9 introduced in [3], where the uniform dependence in* z *is removed. Moreover, one obtains*

$$\langle h(\theta) - h(\theta'), \theta - \theta' \rangle \geq \hat{a}|\theta - \theta'|^2,$$

where $\hat{a} = \hat{a}_1 + \hat{a}_2$*.*

Remark 4.13. *By Remark 4.2 and Remark 4.12, [40, Theorem 2.1.12] shows that*

$$\langle h(\theta) - h(\theta'), \theta - \theta' \rangle \geq \hat{a}^*|\theta - \theta'|^2 + \frac{1}{\hat{a} + L}|h(\theta) - h(\theta')|^2,$$

where $\hat{a}^* = \hat{a}L/(\hat{a} + L)$*.*

Define

$$\bar{\gamma}_{\max} = \min\{1/2(\hat{a} + L), \hat{a}/(4L_1^2\mathbb{E}[K_\rho(Z_0)])\} \tag{4.8}$$

with $\hat{a} = \hat{a}_1 + \hat{a}_2$ given in Remark 4.12.

Remark 4.14. *Throughout this chapter, the constant* $C^* > 0$ *may take different values at different places, but it is always independent of* $n \in \mathbb{N}$, $\beta > 0$ *and* $d \geq 1$*.*

Under the convexity condition of H, the non-asymptotic bound for $W_2(\mathcal{L}(\theta_n^\gamma), \pi_\beta)$ is obtained with the optimal convergence rate $1/2$. The explicit statement is given below.

Theorem 4.15. *Assume **C-1**, **C-2**, **C-3**, **C-5** and **C-6** hold. Then, for any* $n \in \mathbb{N}$, $0 < \gamma < \bar{\gamma}_{\max}$ *given in* (4.8)*, there exist constants* $C_6, C_7, C_8 > 0$ *such that,*

$$W_2(\mathcal{L}(\theta_n^\gamma), \pi_\beta) \leq C_7 e^{-C_6\gamma n} + C_8\sqrt{\gamma},$$

where $C_6 = \hat{a}^*$, $C_7 = O\left(\sqrt{1 + \frac{d}{\beta}}\right)$ *and* $C_8 = O\left(\sqrt{1 + \frac{d}{\beta}}\right)$*. The explicit constants are provided in* (4.39)*. Moreover, if* $\rho = 0$ *in **C-1**, then the result holds for* $\gamma \in \min\{1/2(\hat{a} + L), 1/(6L_1)\}$*.*

By using Theorem 4.15, one can obtain an upper bound for the expected excess risk $\mathbb{E}[\hat{U}(\hat{\theta})] - \inf_{\theta \in \mathbb{R}^d} \hat{U}(\theta)$ in the convex case.

Corollary 4.16. *Assume **C-1**, **C-2**, **C-3**, **C-5** and **C-6** hold. Then, for every* $0 < \gamma \leq \bar{\gamma}_{\max}$ *given in* (4.8)*, there exist constants* $\hat{C}_4, \hat{C}_5, \hat{C}_6, \hat{C}_7 > 0$ *such that the expected excess risk can be estimated as*

$$\mathbb{E}[\hat{U}(\hat{\theta})] - \inf_{\theta \in \mathbb{R}^d} \hat{U}(\theta) \leq \hat{C}_5 e^{-\hat{C}_4\gamma n} + \hat{C}_6\sqrt{\gamma} + \hat{C}_7/\beta,$$

where $\hat{\theta} = \theta_n^\gamma$*, and* $\hat{C}_4 = \hat{a}^*$, $\hat{C}_5 = O\left(1 + \frac{d}{\beta}\right)$, $\hat{C}_6 = O\left(1 + \frac{d}{\beta}\right)$, $\hat{C}_7 = O\left(d \log\left(C^*\left(1 + \frac{\beta}{d}\right)\right)\right)$*. The explicit constants are provided in* (4.41) *and* (4.42)*.*

4.3 Proofs of the main results: nonconvex case

Denote by $\mathscr{F}_t$ the natural filtration of w_t, $t \in \mathbb{R}_+$. It is a classic result that SDE (4.2) has a unique solution adapted to $(\mathscr{F}_t)_{t \in \mathbb{R}_+}$, since h is Lipschitz-continuous by (4.4). In order to obtain the convergence results in Theorem 4.6 and Corollary 4.8, we first introduce some auxiliary processes.

4.3.1 Further notation and introduction of auxiliary processes

Define the Lyapunov function for each $p \geq 1$ by

$$V_p(\theta) := (1 + |\theta|^2)^{p/2}, \ \theta \in \mathbb{R}^d,$$

and similarly $\mathrm{v}_p(x) := (1+x^2)^{p/2}$, for any real $x \geq 0$. Notice that these functions are twice continuously differentiable and

$$\lim_{|\theta|\to\infty} \frac{\nabla V_p(\theta)}{V_p(\theta)} = 0.$$

Let $\mathscr{P}_{V_p}$ denote the set of $\mu \in \mathscr{P}(\mathbb{R}^d)$ satisfying $\int_{\mathbb{R}^d} V_p(\theta)\,\mu(d\theta) < \infty$.

Consider the following auxiliary processes. For each $\gamma > 0$,

$$\hat{Y}_t^\gamma := \hat{Y}_{\gamma t},\ t \in \mathbb{R}_+.$$

Notice that $\tilde{w}_t^\gamma := w_{\gamma t}/\sqrt{\gamma}$, $t \in \mathbb{R}_+$ is also a Brownian motion and

$$d\hat{Y}_t^\gamma = -\gamma h(\hat{Y}_t^\gamma)\,dt + \sqrt{2\beta^{-1}\gamma}d\tilde{w}_t^\gamma,\ \hat{Y}_0^\gamma = \theta_0.$$

Then, $\mathscr{F}_t^\gamma := \mathscr{F}_{\gamma t}$, $t \in \mathbb{R}_+$ is the natural filtration of $\tilde{w}_t^\gamma$, $t \in \mathbb{R}_+$. One notice that $\mathscr{F}_t^\gamma$ is independent of $\mathscr{G}_\infty \vee \sigma(\theta_0)$. Then, define the continuous-time interpolation of the SGLD algorithm (4.1) as

$$d\bar{\theta}_t^\gamma = -\gamma H(\bar{\theta}_{\lfloor t\rfloor}^\gamma, Z_{\lceil t\rceil})\,dt + \sqrt{2\beta^{-1}\gamma}d\tilde{w}_t^\gamma, \tag{4.9}$$

with initial condition $\bar{\theta}_0^\gamma = \theta_0$. In addition, due to the homogeneous nature of the coefficients of equation (4.9), the law of the interpolated process coincides with the law of the SGLD algorithm (4.1) at grid-points, i.e. $\mathcal{L}(\bar{\theta}_n^\gamma) = \mathcal{L}(\theta_n^\gamma)$, for each $n \in \mathbb{N}$.

Furthermore, consider a continuous-time process $\tilde{Y}_t^{s,v,\gamma}$, $t \geq s$, which denotes the solution of the SDE

$$d\tilde{Y}_t^{s,v,\gamma} = -\gamma h(\tilde{Y}_t^{s,v,\gamma})dt + \sqrt{2\beta^{-1}\gamma}d\tilde{w}_t^\gamma.$$

with initial condition $\tilde{Y}_s^{s,v,\gamma} := v$, $v \in \mathbb{R}^d$.

Definition 4.17. Fix $n \in \mathbb{N}$ and define

$$\bar{Y}_t^{\gamma,n} = \tilde{Y}_t^{nT,\bar{\theta}_{nT}^\gamma,\gamma}$$

where $T := \lfloor 1/\gamma \rfloor$.

Intuitively, $\bar{Y}_t^{\gamma,n}$ is a process started from the value of the SGLD process (4.9) at time nT, i.e. $\bar{\theta}_{nT}^\gamma$, and made run until time $t \geq nT$ with the continuous-time Langevin dynamics.

4.3.2 Preliminary estimates

We proceed by establishing the moment bounds of the processes $(\bar{\theta}_t^\gamma)_{t\geq 0}$ and $(\bar{Y}_t^{\gamma,n})_{t\geq 0}$.

Lemma 4.18. *Assume **C-1**, **C-2** and **C-4** hold. For any $0 < \gamma < \gamma_{\max}$ given in (4.5), $n \in \mathbb{N}$, $t \in (n, n+1]$,*

$$\mathbb{E}\left[|\bar{\theta}_t^\gamma|^2\right] \leq (1 - a\gamma(t-n))(1-a\gamma)^n \mathbb{E}\left[|\theta_0|^2\right] + c_1(\gamma_{\max} + a^{-1}),$$

where

$$c_1 = (c_0 + 2d/\beta), \quad c_0 = 8\mathbb{E}\left[K_1^2(Z_0)\right]a^{-1} + 2b + 4\gamma_{\max}L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 4\gamma_{\max}\mathbb{E}\left[F_*^2(Z_0)\right]. \tag{4.10}$$

In addition, $\sup_t \mathbb{E}|\bar{\theta}_t^\gamma|^2 \leq \mathbb{E}\left[|\theta_0|^2\right] + c_1(\gamma_{\max} + a^{-1}) < \infty$. *Similarly, one obtains*

$$\mathbb{E}\left[|\bar{\theta}_t^\gamma|^4\right] \leq (1 - a\gamma(t-n))(1-a\gamma)^n \mathbb{E}|\theta_0|^4 + c_3(\gamma_{\max} + a^{-1}),$$

where

$$c_3 = (1 + a\gamma_{\max})c_2 + 12d^2\beta^{-2}(\gamma_{\max} + 9a^{-1}) \tag{4.11}$$

with c_2 given in (4.16)*. Moreover, this implies* $\sup_t \mathbb{E}|\bar{\theta}_t^\gamma|^4 < \infty$.

Proof. For any $n \in \mathbb{N}$ and $t \in (n, n+1]$, define $\Delta_{n,t} = \bar{\theta}_n^\gamma - \gamma H(\bar{\theta}_n^\gamma, Z_{n+1})(t-n)$. By using (4.9), it is easily seen that for $t \in (n, n+1]$

$$\mathbb{E}\left[|\bar{\theta}_t^\gamma|^2 \,\middle|\, \bar{\theta}_n^\gamma\right] = \mathbb{E}\left[|\Delta_{n,t}|^2 \,\middle|\, \bar{\theta}_n^\gamma\right] + (2\gamma/\beta)d(t-n).$$

Then, by using **C-1**, **C-2**, **C-4** and Remark 4.1, one obtains

$$\begin{aligned}
&\mathbb{E}\left[|\Delta_{n,t}|^2 \,\middle|\, \bar{\theta}^\gamma_n\right] \\
&= |\bar{\theta}^\gamma_n|^2 - 2\gamma(t-n)\mathbb{E}\left[\left\langle \bar{\theta}^\gamma_n, H(\bar{\theta}^\gamma_n, Z_{n+1})\right\rangle \,\middle|\, \bar{\theta}^\gamma_n\right] + \gamma^2(t-n)^2\mathbb{E}\left[|H(\bar{\theta}^\gamma_n, Z_{n+1})|^2 \,\middle|\, \bar{\theta}^\gamma_n\right] \\
&\leq |\bar{\theta}^\gamma_n|^2 - 2\gamma(t-n)\left\langle \bar{\theta}^\gamma_n, \mathbb{E}\left[A(Z_0)\right]\bar{\theta}^\gamma_n\right\rangle + 2\gamma(t-n)b - 2\gamma(t-n)\mathbb{E}\left[\left\langle \bar{\theta}^\gamma_n, G(\bar{\theta}^\gamma_n, Z_{n+1})\right\rangle \,\middle|\, \bar{\theta}^\gamma_n\right] \\
&\quad + \gamma^2(t-n)^2\mathbb{E}\left[\left((1+|Z_{n+1}|)^{\rho+1}(L_1|\bar{\theta}^\gamma_n| + L_2) + F_*(Z_{n+1})\right)^2 \,\middle|\, \bar{\theta}^\gamma_n\right] \\
&\leq (1-2a\gamma(t-n))|\bar{\theta}^\gamma_n|^2 + 2\gamma(t-n)b + 2\gamma(t-n)\mathbb{E}\left[K_1(Z_0)\right]|\bar{\theta}^\gamma_n| \\
&\quad + 2\gamma^2(t-n)^2L_1^2\mathbb{E}\left[K_\rho(Z_0)\right]|\bar{\theta}^\gamma_n|^2 + 4\gamma^2(t-n)^2L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 4\gamma^2(t-n)^2\mathbb{E}\left[F_*^2(Z_0)\right],
\end{aligned}$$

where the last inequality is obtained by using $(a+b)^2 \leq 2a^2 + 2b^2$, for $a, b \geq 0$ twice. For $\gamma < \gamma_{\max}$ with $\gamma_{\max}$ given in (4.5),

$$\begin{aligned}
\mathbb{E}\left[|\Delta_{n,t}|^2 \,\middle|\, \bar{\theta}^\gamma_n\right] \leq{}& \left(1 - \frac{3}{2}a\gamma(t-n)\right)|\bar{\theta}^\gamma_n|^2 + 2\gamma(t-n)\mathbb{E}\left[K_1(Z_0)\right]|\bar{\theta}^\gamma_n| \\
&+ 2\gamma(t-n)b + 4\gamma^2(t-n)^2L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 4\gamma^2(t-n)^2\mathbb{E}\left[F_*^2(Z_0)\right].
\end{aligned}$$

For $|\bar{\theta}^\gamma_n| > 4\mathbb{E}\left[K_1(Z_0)\right]a^{-1}$, one obtains

$$-\frac{1}{2}a\gamma(t-n)|\bar{\theta}^\gamma_n|^2 + 2\gamma(t-n)\mathbb{E}\left[K_1(Z_0)\right]|\bar{\theta}^\gamma_n| < 0,$$

which implies

$$\begin{aligned}
\mathbb{E}\left[|\Delta_{n,t}|^2 \,\middle|\, \bar{\theta}^\gamma_n\right] \leq{}& (1 - a\gamma(t-n))|\bar{\theta}^\gamma_n|^2 + 2\gamma(t-n)b \\
&+ 4\gamma^2(t-n)^2L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 4\gamma^2(t-n)^2\mathbb{E}\left[F_*^2(Z_0)\right].
\end{aligned}$$

For $|\bar{\theta}^\gamma_n| \leq 4\mathbb{E}\left[K_1(Z_0)\right]a^{-1}$, we have

$$\begin{aligned}
\mathbb{E}\left[|\Delta_{n,t}|^2 \,\middle|\, \bar{\theta}^\gamma_n\right] \leq{}& \left(1 - \frac{3}{2}a\gamma(t-n)\right)|\bar{\theta}^\gamma_n|^2 + 8\gamma(t-n)\mathbb{E}\left[K_1^2(Z_0)\right]a^{-1} \\
&+ 2\gamma(t-n)b + 4\gamma^2(t-n)^2L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 4\gamma^2(t-n)^2\mathbb{E}\left[F_*^2(Z_0)\right].
\end{aligned}$$

Combining the two cases yields

$$\mathbb{E}\left[|\Delta_{n,t}|^2 \,\middle|\, \bar{\theta}^\gamma_n\right] \leq (1 - a\gamma(t-n))|\bar{\theta}^\gamma_n|^2 + \gamma(t-n)c_0,$$

where $c_0 = 8\mathbb{E}\left[K_1^2(Z_0)\right]a^{-1} + 2b + 4\gamma_{\max}L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 4\gamma_{\max}\mathbb{E}\left[F_*^2(Z_0)\right]$. Therefore, one obtains

$$\mathbb{E}\left[|\bar{\theta}^\gamma_t|^2 \,\middle|\, \bar{\theta}^\gamma_n\right] \leq (1 - a\gamma(t-n))|\bar{\theta}^\gamma_n|^2 + \gamma(t-n)c_1,$$

where $c_1 = (c_0 + 2d/\beta)$ and the result follows by induction. To calculate a higher moment, denote by $\Xi^\gamma_{n,t} = \{2\gamma\beta^{-1}\}^{1/2}(\tilde{w}^\gamma_t - \tilde{w}^\gamma_n)$, for $t \in (n, n+1]$, one calculates

$$\begin{aligned}
\mathbb{E}\left[|\bar{\theta}^\gamma_t|^4 \,\middle|\, \bar{\theta}^\gamma_n\right] &= \mathbb{E}\left[\left(|\Delta_{n,t}|^2 + |\Xi^\gamma_{n,t}|^2 + 2\left\langle \Delta_{n,t}, \Xi^\gamma_{n,t}\right\rangle\right)^2 \,\middle|\, \bar{\theta}^\gamma_n\right] \\
&= \mathbb{E}\left[|\Delta_{n,t}|^4 + |\Xi^\gamma_{n,t}|^4 + 2|\Delta_{n,t}|^2|\Xi^\gamma_{n,t}|^2 + 4|\Delta_{n,t}|^2\left\langle \Delta_{n,t}, \Xi^\gamma_{n,t}\right\rangle\right. \\
&\qquad \left. + 4|\Xi^\gamma_{n,t}|^2\left\langle \Delta_{n,t}, \Xi^\gamma_{n,t}\right\rangle + 4\left(\left\langle \Delta_{n,t}, \Xi^\gamma_{n,t}\right\rangle\right)^2 \,\middle|\, \bar{\theta}^\gamma_n\right] \\
&\leq \mathbb{E}\left[|\Delta_{n,t}|^4 + |\Xi^\gamma_{n,t}|^4 + 6|\Delta_{n,t}|^2|\Xi^\gamma_{n,t}|^2 \,\middle|\, \bar{\theta}^\gamma_n\right] \\
&\leq (1 + a\gamma(t-n))\mathbb{E}\left[|\Delta_{n,t}|^4 \,\middle|\, \bar{\theta}^\gamma_n\right] + (1 + 9/(a\gamma(t-n)))\mathbb{E}\left[|\Xi^\gamma_{n,t}|^4\right]. \qquad (4.12)
\end{aligned}$$

where the last inequality holds due to $2ab \leq \varepsilon a^2 + \varepsilon^{-1}b^2$, for $a, b \geq 0$ and $\varepsilon > 0$ with $\varepsilon = a\gamma(t-n)$. Then, one continues with calculating

$$\begin{aligned}
&\mathbb{E}\left[|\Delta_{n,t}|^4 \,\middle|\, \bar{\theta}^\gamma_n\right] \\
&= \mathbb{E}\left[\left(|\bar{\theta}^\gamma_n|^2 - 2\gamma(t-n)\left\langle \bar{\theta}^\gamma_n, H(\bar{\theta}^\gamma_n, Z_{n+1})\right\rangle + \gamma^2(t-n)^2|H(\bar{\theta}^\gamma_n, Z_{n+1})|^2\right)^2 \,\middle|\, \bar{\theta}^\gamma_n\right]
\end{aligned}$$

$$\leq |\bar{\theta}^\gamma_n|^4 + \mathbb{E}\left[6\gamma^2(t-n)^2|\bar{\theta}^\gamma_n|^2|H(\bar{\theta}^\gamma_n, Z_{n+1})|^2 - 4\gamma(t-n)\left\langle \bar{\theta}^\gamma_n, H(\bar{\theta}^\gamma_n, Z_{n+1})\right\rangle |\bar{\theta}^\gamma_n|^2 \right.$$
$$\left. -4\gamma^3(t-n)^3|H(\bar{\theta}^\gamma_n, Z_{n+1})|^2\left\langle \bar{\theta}^\gamma_n, H(\bar{\theta}^\gamma_n, Z_{n+1})\right\rangle + \gamma^4(t-n)^4|H(\bar{\theta}^\gamma_n, Z_{n+1})|^4 \,\middle|\, \bar{\theta}^\gamma_n\right].$$

By Remark 4.1, for $q \geq 1$, one observes

$$\mathbb{E}\left[|H(\bar{\theta}^\gamma_n, Z_{n+1})|^q \,\middle|\, \bar{\theta}^\gamma_n\right] \leq \mathbb{E}\left[(1+|Z_0|)^{q\rho+q}\right](2^{q-1}L_1^q|\bar{\theta}^\gamma_n|^q + 2^{2q-2}L_2^q) + 2^{2q-2}\mathbb{E}\left[F_*^q(Z_0)\right]. \quad (4.13)$$

Then, by using **C-4** and by taking $q = 2, 3, 4$ in (4.13), one obtains

$$\begin{aligned}
&\mathbb{E}\left[|\Delta_{n,t}|^4 \,\middle|\, \bar{\theta}^\gamma_n\right] \\
&\leq (1-4a\gamma(t-n))|\bar{\theta}^\gamma_n|^4 + 4b\gamma(t-n)|\bar{\theta}^\gamma_n|^2 + 4\gamma(t-n)\mathbb{E}\left[K_1(Z_0)\right]|\bar{\theta}^\gamma_n|^3 \\
&\quad + 12\gamma^2(t-n)^2L_1^2\mathbb{E}\left[K_\rho(Z_0)\right]|\bar{\theta}^\gamma_n|^4 + 24\gamma^2(t-n)^2\left(L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^2(Z_0)\right]\right)|\bar{\theta}^\gamma_n|^2 \\
&\quad + 16\gamma^3(t-n)^3L_1^3\mathbb{E}\left[K_\rho(Z_0)\right]|\bar{\theta}^\gamma_n|^4 + 64\gamma^3(t-n)^3\left(L_2^3\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^3(Z_0)\right]\right)|\bar{\theta}^\gamma_n| \\
&\quad + 8\gamma^4(t-n)^4L_1^4\mathbb{E}\left[K_\rho(Z_0)\right]|\bar{\theta}^\gamma_n|^4 + 64\gamma^4(t-n)^4\left(L_2^4\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^4(Z_0)\right]\right),
\end{aligned}$$

which implies, by using $\gamma < \gamma_{\max}$

$$\begin{aligned}
\mathbb{E}\left[|\Delta_{n,t}|^4 \,\middle|\, \bar{\theta}^\gamma_n\right] &\leq (1-3a\gamma(t-n))|\bar{\theta}^\gamma_n|^4 + 4\gamma(t-n)\mathbb{E}\left[K_1(Z_0)\right]|\bar{\theta}^\gamma_n|^3 \\
&\quad + 4b\gamma(t-n)|\bar{\theta}^\gamma_n|^2 + 24\gamma^2(t-n)^2\left(L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^2(Z_0)\right]\right)|\bar{\theta}^\gamma_n|^2 \\
&\quad + 64\gamma^3(t-n)^3\left(L_2^3\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^3(Z_0)\right]\right)|\bar{\theta}^\gamma_n| \\
&\quad + 64\gamma^4(t-n)^4\left(L_2^4\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^4(Z_0)\right]\right).
\end{aligned}$$

For $|\bar{\theta}^\gamma_n| > 12\mathbb{E}\left[K_1(Z_0)\right]a^{-1}$, one obtains

$$-\frac{a\gamma(t-n)}{3}|\bar{\theta}^\gamma_n|^4 + 4\gamma(t-n)\mathbb{E}\left[K_1(Z_0)\right]|\bar{\theta}^\gamma_n|^3 < 0,$$

similarly, for $|\bar{\theta}^\gamma_n| > (12ba^{-1} + 72a^{-1}\gamma_{\max}\left(L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^2(Z_0)\right]\right))^{1/2}$, we have

$$-\frac{a\gamma(t-n)}{3}|\bar{\theta}^\gamma_n|^4 + 4b\gamma(t-n)|\bar{\theta}^\gamma_n|^2 + 24\gamma^2(t-n)^2\left(L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^2(Z_0)\right]\right)|\bar{\theta}^\gamma_n|^2 < 0,$$

moreover, for $|\bar{\theta}^\gamma_n| > (192a^{-1}\gamma_{\max}^2\left(L_2^3\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^3(Z_0)\right]\right))^{1/3}$

$$-\frac{a\gamma(t-n)}{3}|\bar{\theta}^\gamma_n|^4 + 64\gamma^3(t-n)^3\left(L_2^3\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^3(Z_0)\right]\right)|\bar{\theta}^\gamma_n| < 0.$$

Denote by

$$\begin{aligned}
M = \max\Big\{&12\mathbb{E}\left[K_1(Z_0)\right]a^{-1}, (12ba^{-1} + 72a^{-1}\gamma_{\max}\left(L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^2(Z_0)\right]\right))^{1/2}, \\
&(192a^{-1}\gamma_{\max}^2\left(L_2^3\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^3(Z_0)\right]\right))^{1/3}\Big\}.
\end{aligned} \quad (4.14)$$

For $|\bar{\theta}^\gamma_n| > M$, one obtains

$$\mathbb{E}\left[|\Delta_{n,t}|^4 \,\middle|\, \bar{\theta}^\gamma_n\right] \leq (1-2a\gamma(t-n))|\bar{\theta}^\gamma_n|^4 + 64\gamma^4(t-n)^4\left(L_2^4\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^4(Z_0)\right]\right).$$

As for $|\bar{\theta}^\gamma_n| \leq M$, we have

$$\begin{aligned}
\mathbb{E}\left[|\Delta_{n,t}|^4 \,\middle|\, \bar{\theta}^\gamma_n\right] &\leq (1-3a\gamma(t-n))|\bar{\theta}^\gamma_n|^4 + 4\gamma(t-n)\mathbb{E}\left[K_1(Z_0)\right]M^3 + 4b\gamma(t-n)M^2 \\
&\quad + 24\gamma^2(t-n)^2\left(L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^2(Z_0)\right]\right)M^2 \\
&\quad + 64\gamma^3(t-n)^3\left(L_2^3\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^3(Z_0)\right]\right)M \\
&\quad + 64\gamma^4(t-n)^4\left(L_2^4\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^4(Z_0)\right]\right).
\end{aligned}$$

Combining the two cases yields

$$\mathbb{E}\left[|\Delta_{n,t}|^4 \,\middle|\, \bar{\theta}_n^\gamma\right] \le (1 - 2a\gamma(t-n))|\bar{\theta}_n^\gamma|^4 + \gamma(t-n)c_2, \tag{4.15}$$

where

$$c_2 = 4\mathbb{E}\left[K_1(Z_0)\right] M^3 + 4bM^2 + 152(1+\gamma_{\max})^3 \left((1+L_2)^4\mathbb{E}\left[K_\rho(Z_0)\right] + (1+\mathbb{E}\left[F_*^4(Z_0)\right])\right)(1+M)^2 \tag{4.16}$$

with M given in (4.14). Substituting (4.15) into (4.12), one obtains

$$\begin{aligned}
\mathbb{E}\left[|\bar{\theta}_t^\gamma|^4 \,\middle|\, \bar{\theta}_n^\gamma\right] &\le (1 + a\gamma(t-n))(1 - 2a\gamma(t-n))|\bar{\theta}_n^\gamma|^4 \\
&\quad + (1 + a\gamma(t-n))\gamma(t-n)c_2 + 12d^2\gamma^2\beta^{-2}(t-n)^2(1 + 9/(a\gamma(t-n))) \\
&\le (1 - a\gamma(t-n))|\bar{\theta}_n^\gamma|^4 + \gamma(t-n)c_3,
\end{aligned}$$

where $c_3 = (1 + a\gamma_{\max})c_2 + 12d^2\beta^{-2}(\gamma_{\max} + 9a^{-1})$. The proof completes by induction. □

Remark 4.19. *One notices that in Lemma 4.18, the step-size restriction is the following:*

$$\hat{\gamma}_{\max} = \min\left\{\frac{a}{24L_1^2\mathbb{E}\left[K_\rho(Z_0)\right]}, \frac{a^{1/2}}{8(L_1^3\mathbb{E}\left[K_\rho(Z_0)\right])^{1/2}}, \frac{a^{1/3}}{(32L_1^4\mathbb{E}\left[K_\rho(Z_0)\right])^{1/3}}, \frac{1}{4a}\right\}.$$

Theorem 4.6 and Corollary 4.8 still hold by using $\hat{\gamma}_{\max}$. However, in order to make notation compact, the restriction is chosen to be $\gamma_{\max}$ given in (4.5)*, which can be deduced from the above expression.*

Corollary 4.20. *Assume **C-1**, **C-2** and **C-4** hold. For any $0 < \gamma < \gamma_{\max}$ given in* (4.5)*, $n \in \mathbb{N}$, $t \in (n, n+1]$,*

$$\mathbb{E}[V_4(\bar{\theta}_t^\gamma)] \le 2(1 - a\gamma)^{\lfloor t \rfloor}\mathbb{E}[V_4(\theta_0)] + 2c_3(\gamma_{\max} + a^{-1}) + 2,$$

where c_3 is given in (4.11)*.*

Next, we present a drift condition associated with the SDE (4.2), which will be used to obtain the moment bounds of the process $(\bar{Y}_t^{\gamma,n})_{t\ge0}$.

Lemma 4.21. *Assume **C-1**, **C-2** and **C-4** hold. Then, for each $p \ge 2$, $\theta \in \mathbb{R}^d$,*

$$\frac{\Delta V_p}{\beta} - \langle h(\theta), \nabla V_p(\theta)\rangle \le -\bar{c}(p)V_p(\theta) + \tilde{c}(p),$$

where $\bar{c}(p) = ap/4$ and $\tilde{c}(p) = (3/4)ap\mathrm{v}_p(\overline{M}_p)$ with $\overline{M}_p$ given in (4.17)*.*

Proof. One notices that, by **C-1** and **C-2**, for any $\theta \in \mathbb{R}^d$, $h(\theta) = \mathbb{E}[H(\theta, Z_0)] = \mathbb{E}[F(\theta, Z_0) + G(\theta, Z_0)]$. Then, one calculates,

$$\begin{aligned}
&\frac{\Delta V_p}{\beta} - \langle h(\theta), \nabla V_p(\theta)\rangle \\
&= \beta^{-1}p(p-2)|\theta|^2V_{p-4}(\theta) + \beta^{-1}pdV_{p-2}(\theta) - pV_{p-2}(\theta)\langle\mathbb{E}[F(\theta, Z_0) + G(\theta, Z_0)], \theta\rangle \\
&\le -apV_p(\theta) + (ap + bp + \beta^{-1}p(p-2) + \beta^{-1}pd)V_{p-2}(\theta) + p\mathbb{E}\left[K_1(Z_0)\right]|\theta|V_{p-2}(\theta),
\end{aligned}$$

where the last inequality is obtained due to **C-4**. Denote by

$$\overline{M}_p = \sqrt{(4/3 + 4b/(3a) + 4d/(3a\beta) + 4(p-2)/(3a\beta) + 4\mathbb{E}\left[K_1(Z_0)\right]/(3a))^2 - 1}. \tag{4.17}$$

For $|\theta| > \overline{M}_p$, by observing $|\theta| \le \sqrt{1 + |\theta|^2}$, one obtains $\frac{\Delta V_p}{\beta} - \langle h(\theta), \nabla V_p(\theta)\rangle \le -(ap/4)V_p(\theta)$, whereas for $|\theta| \le \overline{M}_p$, one observes that $\frac{\Delta V_p}{\beta} - \langle h(\theta), \nabla V_p(\theta)\rangle \le (3/4)ap\mathrm{v}_p(\overline{M}_p)$. Combining the two cases yields the desired result. □

The following Lemma provides upper bounds for the second and the fourth moment of the process $(\bar{Y}_t^{\gamma,n})_{t\ge0}$.

Lemma 4.22. *Assume **C-1**, **C-2** and **C-4** hold. For any* $0 < \gamma < \gamma_{\max}$ *given in* (4.5), $t \geq nT$, $n \in \mathbb{N}$, *one obtains the following inequality*

$$\mathbb{E}[V_2(\bar{Y}_t^{\gamma,n})] \leq e^{-a\gamma t/2}\mathbb{E}[V_2(\theta_0)] + 3\mathrm{v}_2(\overline{M}_2) + c_1(\gamma_{\max} + a^{-1}) + 1,$$

where the process $\bar{Y}_t^{\gamma,n}$ *is defined in Definition 4.17 and* c_1 *is given in* (4.10). *Furthermore,*

$$\mathbb{E}[V_4(\bar{Y}_t^{\gamma,n})] \leq 2e^{-a\gamma t}\mathbb{E}[V_4(\theta_0)] + 3\mathrm{v}_4(\overline{M}_4) + 2c_3(\gamma_{\max} + a^{-1}) + 2,$$

where c_3 *is given in* (4.11).

Proof. For any $p \geq 1$, application of Ito's lemma and taking expectation yields

$$\mathbb{E}[V_p(\bar{Y}_t^{\gamma,n})] = \mathbb{E}[V_p(\bar{\theta}_{nT}^{\gamma})] + \int_{nT}^{t} \mathbb{E}\left[\gamma \frac{\Delta V_p(\bar{Y}_s^{\gamma,n})}{\beta} - \gamma \langle h(\bar{Y}_s^{\gamma,n}), \nabla V_p(\bar{Y}_s^{\gamma,n})\rangle\right] ds.$$

Differentiating both sides and using Lemma 4.21, we arrive at

$$\frac{d}{dt}\mathbb{E}[V_p(\bar{Y}_t^{\gamma,n})] = \mathbb{E}\left[\gamma \frac{\Delta V_p(\bar{Y}_t^{\gamma,n})}{\beta} - \gamma \langle h(\bar{Y}_t^{\gamma,n}), \nabla V_p(\bar{Y}_t^{\gamma,n})\rangle\right] \leq -\gamma \bar{c}(p)\mathbb{E}[V_p(\bar{Y}_t^{\gamma,n})] + \gamma \tilde{c}(p),$$

which yields

$$\begin{aligned}\mathbb{E}[V_p(\bar{Y}_t^{\gamma,n})] &\leq e^{-\gamma(t-nT)\bar{c}(p)}\mathbb{E}[V_p(\bar{\theta}_{nT}^{\gamma})] + \frac{\tilde{c}(p)}{\bar{c}(p)}\left(1 - e^{-\gamma\bar{c}(p)(t-nT)}\right) \\ &\leq e^{-\gamma(t-nT)\bar{c}(p)}\mathbb{E}[V_p(\bar{\theta}_{nT}^{\gamma})] + \frac{\tilde{c}(p)}{\bar{c}(p)}.\end{aligned}$$

Now for $p = 2$,one obtains

$$\begin{aligned}\mathbb{E}[V_2(\bar{Y}_t^{\gamma,n})] &\leq e^{-\gamma(t-nT)\bar{c}(2)}\mathbb{E}[V_2(\bar{\theta}_{nT}^{\gamma})] + \frac{\tilde{c}(2)}{\bar{c}(2)} \\ &\leq (1 - a\gamma)^{nT}e^{-\gamma(t-nT)\bar{c}(2)}\mathbb{E}[V_2(\theta_0)] + \frac{\tilde{c}(2)}{\bar{c}(2)} + c_1(\gamma_{\max} + a^{-1}) + 1 \\ &\leq e^{-a\gamma t/2}\mathbb{E}[V_2(\theta_0)] + 3\mathrm{v}_2(\overline{M}_2) + c_1(\gamma_{\max} + a^{-1}) + 1,\end{aligned}$$

where the last inequality holds due to $1 - z \leq e^{-z}$ for $z \geq 0$ and $\bar{c}(2) = a/2$. Similarly, for $p = 4$, by using Corollary 4.20 one obtains

$$\begin{aligned}\mathbb{E}[V_4(\bar{Y}_t^{\gamma,n})] &\leq e^{-\gamma(t-nT)\bar{c}(4)}\mathbb{E}[V_4(\bar{\theta}_{nT}^{\gamma})] + \frac{\tilde{c}(4)}{\bar{c}(4)} \\ &\leq 2(1 - a\gamma)^{nT}e^{-\gamma(t-nT)\bar{c}(4)}\mathbb{E}[V_4(\theta_0)] + \frac{\tilde{c}(4)}{\bar{c}(4)} + 2c_3(\gamma_{\max} + a^{-1}) + 2 \\ &\leq 2e^{-a\gamma t}\mathbb{E}[V_4(\theta_0)] + 3\mathrm{v}_4(\overline{M}_4) + 2c_3(\gamma_{\max} + a^{-1}) + 2,\end{aligned}$$

where the last inequality holds due to $1 - z \leq e^{-z}$ for $z \geq 0$ and $\bar{c}(4) = a$. □

4.3.3 Proof of the main theorems

We first introduce a functional which is crucial for obtaining the convergence rate in W_1. For any $p \geq 1$, $\mu, \nu \in \mathscr{P}_{V_p}$,

$$\tilde{W}_{1,p}(\mu,\nu) := \inf_{\zeta \in \mathscr{C}(\mu,\nu)} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} [1 \wedge |\theta - \theta'|](1 + V_p(\theta) + V_p(\theta'))\zeta(d\theta d\theta'), \tag{4.18}$$

and it satisfies trivially

$$W_1(\mu,\nu) \leq \tilde{W}_{1,p}(\mu,\nu). \tag{4.19}$$

The case $p = 2$, i.e. $\tilde{W}_{1,2}$, is used throughout the section. The result below states a contraction property of $\tilde{W}_{1,2}$.

Proposition 4.23. *Let $\hat{Y}'_t$, $t \in \mathbb{R}_+$ be the solution of* (4.2) *with initial condition $\hat{Y}'_0 = \theta'_0$ which is independent of $\mathscr{F}_\infty$ and satisfies $|\theta_0| \in \mathscr{L}^2$. Then,*

$$\tilde{W}_{1,2}(\mathcal{L}(\hat{Y}_t), \mathcal{L}(\hat{Y}'_t)) \leq \hat{c} e^{-\dot{c}t} \tilde{W}_{1,2}(\mathcal{L}(\theta_0), \mathcal{L}(\theta'_0)),$$

where the constants $\dot{c}$ and $\hat{c}$ are given in Lemma 4.28.

Proof. See Proposition 3.14 of [9]. □

We aim to establish the non-asymptotic bound between $\mathcal{L}(\bar{\theta}^\gamma_t)$ and $\mathcal{L}(\hat{Y}^\gamma_t)$, $t \in [nT, (n+1)T]$, in Wasserstein-1 distance. To achieve this, we consider the following decomposition using the auxiliary process $\bar{Y}^{\gamma,n}_t$ introduced in Definition 4.17:

$$W_1(\mathcal{L}(\bar{\theta}^\gamma_t), \mathcal{L}(\hat{Y}^\gamma_t)) \leq W_1(\mathcal{L}(\bar{\theta}^\gamma_t), \mathcal{L}(\bar{Y}^{\gamma,n}_t)) + W_1(\mathcal{L}(\bar{Y}^{\gamma,n}_t), \mathcal{L}(\hat{Y}^\gamma_t)). \tag{4.20}$$

One notices that when $1 < \gamma \leq \gamma_{\max}$, the result holds trivially. Thus, we consider the case $0 < \gamma \leq 1$, which implies $1/2 < \gamma T \leq 1$.

An upper bound for the first term in (4.20) is obtained in the Lemma below.

Lemma 4.24. *Assume **C-1, C-2, C-3** and **C-4** hold. For any $0 < \gamma < \gamma_{\max}$ given in* (4.5)*, $t \in [nT, (n+1)T]$,*

$$W_1(\mathcal{L}(\bar{\theta}^\gamma_t), \mathcal{L}(\bar{Y}^{\gamma,n}_t)) \leq \sqrt{\gamma}(e^{-an/2}\bar{C}_{2,1}\mathbb{E}[V_2(\theta_0)] + \bar{C}_{2,2})^{1/2},$$

where $\bar{C}_{2,1}$ and $\bar{C}_{2,2}$ are given in (4.23).

Proof. To handle the first term in (4.20), we start by establishing an upper bound in Wasserstein-2 distance and the statement follows by noticing $W_1 \leq W_2$. By employing synchronous coupling, using (4.9) and the definition of $\bar{Y}^{\gamma,n}_t$ in Definition 4.17, one obtains

$$|\bar{Y}^{\gamma,n}_t - \bar{\theta}^\gamma_t| \leq \gamma \left| \int_{nT}^t \left[H(\bar{\theta}^\gamma_{\lfloor s \rfloor}, Z_{\lceil s \rceil}) - h(\bar{Y}^{\gamma,n}_s) \right] ds \right|.$$

Then, the triangle inequality leads

$$|\bar{Y}^{\gamma,n}_t - \bar{\theta}^\gamma_t| \leq \gamma \left| \int_{nT}^t \left[H(\bar{\theta}^\gamma_{\lfloor s \rfloor}, Z_{\lceil s \rceil}) - h(\bar{\theta}^\gamma_{\lfloor s \rfloor}) \right] ds \right| + \gamma \left| \int_{nT}^t \left[h(\bar{\theta}^\gamma_{\lfloor s \rfloor}) - h(\bar{Y}^{\gamma,n}_s) \right] ds \right|.$$

Taking squares on both sides and the application of Remark 4.2 yield

$$|\bar{Y}^{\gamma,n}_t - \bar{\theta}^\gamma_t|^2 \leq 2\gamma^2 \left| \int_{nT}^t \left[H(\bar{\theta}^\gamma_{\lfloor s \rfloor}, Z_{\lceil s \rceil}) - h(\bar{\theta}^\gamma_{\lfloor s \rfloor}) \right] ds \right|^2 + 2\gamma L^2 \int_{nT}^t \left| \bar{\theta}^\gamma_{\lfloor s \rfloor} - \bar{Y}^{\gamma,n}_s \right|^2 ds.$$

By taking expectations on both sides and by using $(a+b)^2 \leq 2a^2 + 2b^2$, for $a, b > 0$, one obtains

$$\begin{aligned} \mathbb{E}\left[|\bar{Y}^{\gamma,n}_t - \bar{\theta}^\gamma_t|^2\right] &\leq 2\gamma^2 \mathbb{E}\left[\left| \int_{nT}^t \left[H(\bar{\theta}^\gamma_{\lfloor s \rfloor}, Z_{\lceil s \rceil}) - h(\bar{\theta}^\gamma_{\lfloor s \rfloor}) \right] ds \right|^2 \right] \\ &\quad + 4\gamma L^2 \int_{nT}^t \mathbb{E}\left[\left| \bar{\theta}^\gamma_{\lfloor s \rfloor} - \bar{\theta}^\gamma_s \right|^2 \right] ds + 4\gamma L^2 \int_{nT}^t \mathbb{E}\left[|\bar{\theta}^\gamma_s - \bar{Y}^{\gamma,n}_s|^2 \right] ds, \end{aligned}$$

which implies due to $\gamma T \leq 1$ and Lemma C.2

$$\begin{aligned} \mathbb{E}\left[|\bar{Y}^{\gamma,n}_t - \bar{\theta}^\gamma_t|^2\right] &\leq 4\gamma L^2 (e^{-a\gamma nT} \bar{\sigma}_Y \mathbb{E}[V_2(\theta_0)] + \tilde{\sigma}_Y) + 4\gamma L^2 \int_{nT}^t \mathbb{E}\left[|\bar{\theta}^\gamma_s - \bar{Y}^{\gamma,n}_s|^2 \right] ds \\ &\quad + 2\gamma^2 \mathbb{E}\left[\left| \int_{nT}^t \left[H(\bar{\theta}^\gamma_{\lfloor s \rfloor}, Z_{\lceil s \rceil}) - h(\bar{\theta}^\gamma_{\lfloor s \rfloor}) \right] ds \right|^2 \right], \end{aligned} \tag{4.21}$$

where $\bar{\sigma}_Y$ and $\tilde{\sigma}_Y$ are provided in (C.5). Next, we bound the last term in (4.21) by partitioning the integral. For $K \in \mathbb{N}$, assume that $nT + K \leq t \leq nT + K + 1$ where $K + 1 \leq T$. Thus we can write

$$\left| \int_{nT}^{t} \left[H(\bar{\theta}^{\gamma}_{\lfloor s \rfloor}, Z_{\lceil s \rceil}) - h(\bar{\theta}^{\gamma}_{\lfloor s \rfloor}) \right] ds \right| = \left| \sum_{k=1}^{K} I_k + R_K \right|$$

where

$$\begin{aligned} I_k &= H(\bar{\theta}^{\gamma}_{nT+k-1}, Z_{nT+k}) - h(\bar{\theta}^{\gamma}_{nT+k-1}), \\ R_K &= (t - (nT + K))(H(\bar{\theta}^{\gamma}_{nT+K}, Z_{nT+K+1}) - h(\bar{\theta}^{\gamma}_{nT+K})). \end{aligned}$$

Taking squares of both sides

$$\left| \sum_{k=1}^{K} I_k + R_K \right|^2 = \sum_{k=1}^{K} |I_k|^2 + 2 \sum_{k=2}^{K} \sum_{j=1}^{k-1} \langle I_k, I_j \rangle + 2 \sum_{k=1}^{K} \langle I_k, R_K \rangle + |R_K|^2,$$

Finally, we take expectations of both sides. Define the filtration $\mathscr{H}_t = \mathscr{F}^{\gamma}_{\infty} \vee \mathscr{G}_{\lfloor t \rfloor}$. We first note that for any $k = 2, \dots, K$, $j = 1, \dots, k-1$,

$$\begin{aligned} &\mathbb{E}\langle I_k, I_j \rangle \\ &= \mathbb{E}\left[\mathbb{E}[\langle I_k, I_j \rangle | \mathscr{H}_{nT+k-1}] \right], \\ &= \mathbb{E}\left[\mathbb{E}\left[\left\langle H(\bar{\theta}^{\gamma}_{nT+k-1}, Z_{nT+k}) - h(\bar{\theta}^{\gamma}_{nT+k-1}), H(\bar{\theta}^{\gamma}_{nT+j-1}, Z_{nT+j}) - h(\bar{\theta}^{\gamma}_{nT+j-1}) \right\rangle \middle| \mathscr{H}_{nT+k-1} \right] \right], \\ &= \mathbb{E}\left[\left\langle \mathbb{E}\left[H(\bar{\theta}^{\gamma}_{nT+k-1}, Z_{nT+k}) - h(\bar{\theta}^{\gamma}_{nT+k-1}) \middle| \mathscr{H}_{nT+k-1} \right], H(\bar{\theta}^{\gamma}_{nT+j-1}, Z_{nT+j}) - h(\bar{\theta}^{\gamma}_{nT+j-1}) \right\rangle \right], \\ &= 0. \end{aligned}$$

By the same argument, $\mathbb{E}\langle I_k, R_K \rangle = 0$ for all $1 \leq k \leq K$. Therefore, the upper bound of the last term in (4.21) is given by

$$\begin{aligned} 2\gamma^2 \mathbb{E}\left[\left| \int_{nT}^{t} \left[H(\bar{\theta}^{\gamma}_{\lfloor s \rfloor}, Z_{\lceil s \rceil}) - h(\bar{\theta}^{\gamma}_{\lfloor s \rfloor}) \right] ds \right|^2 \right] &= 2\gamma^2 \sum_{k=1}^{K} \mathbb{E}\left[|I_k|^2 \right] + 2\gamma^2 \mathbb{E}\left[|R_K|^2 \right] \\ &\leq 2\gamma (e^{-a\gamma nT} \bar{\sigma}_Z \mathbb{E}[V_2(\theta_0)] + \tilde{\sigma}_Z), \end{aligned}$$

where the inequality holds due to Lemma C.1 and $\bar{\sigma}_Z$ and $\tilde{\sigma}_Z$ are provided in (C.4). Substituting the above results into (4.21), one obtains

$$\begin{aligned} \mathbb{E}\left[\left| \bar{Y}^{\gamma,n}_t - \bar{\theta}^{\gamma}_t \right|^2 \right] &\leq 4\gamma L^2 \int_{nT}^{t} \mathbb{E}\left[\left| \bar{\theta}^{\gamma}_s - \bar{Y}^{\gamma,n}_s \right|^2 \right] ds \\ &\quad + 4\gamma e^{-a\gamma nT} (L^2 \bar{\sigma}_Y + \bar{\sigma}_Z) \mathbb{E}[V_2(\theta_0)] + 4\gamma (L^2 \tilde{\sigma}_Y + \tilde{\sigma}_Z), \end{aligned}$$

Using Grönwall's inequality yields

$$\mathbb{E}\left[\left| \bar{Y}^{\gamma,n}_t - \bar{\theta}^{\gamma}_t \right|^2 \right] \leq 4\gamma e^{4L^2} \left[e^{-a\gamma nT} (L^2 \bar{\sigma}_Y + \bar{\sigma}_Z) \mathbb{E}[V_2(\theta_0)] + (L^2 \tilde{\sigma}_Y + \tilde{\sigma}_Z) \right],$$

which implies by $\gamma T \geq 1/2$,

$$W_2^2(\mathcal{L}(\bar{\theta}^{\gamma}_t), \mathcal{L}(\bar{Y}^{\gamma,n}_t)) \leq \mathbb{E}\left[\left| \bar{Y}^{\gamma,n}_t - \bar{\theta}^{\gamma}_t \right|^2 \right] \leq \gamma (e^{-an/2} \bar{C}_{2,1} \mathbb{E}[V_2(\theta_0)] + \bar{C}_{2,2}), \tag{4.22}$$

where

$$\bar{C}_{2,1} = 4e^{4L^2} (L^2 \bar{\sigma}_Y + \bar{\sigma}_Z), \quad \bar{C}_{2,2} = 4e^{4L^2} (L^2 \tilde{\sigma}_Y + \tilde{\sigma}_Z) \tag{4.23}$$

with $\bar{\sigma}_Y$, $\tilde{\sigma}_Y$ provided in (C.5) and $\bar{\sigma}_Z$, $\tilde{\sigma}_Z$ given in (C.4). □

Then, the following Lemma provides the bound for the second term in (4.20).

Lemma 4.25. *Assume **C-1**, **C-2**, **C-3** and **C-4** hold. For any* $0 < \gamma < \gamma_{\max}$ *given in* (4.5), $t \in [nT, (n+1)T]$,

$$W_1(\mathcal{L}(\bar{Y}_t^{\gamma,n}), \mathcal{L}(\hat{Y}_t^{\gamma})) \leq \sqrt{\gamma}(e^{-\min\{\dot{c},a/2\}n/2}\bar{C}_{2,3}\mathbb{E}[V_4(\theta_0)] + \bar{C}_{2,4}),$$

where $\bar{C}_{2,3}$, $\bar{C}_{2,4}$ *is given in* (4.24).

Proof. To upper bound the second term $W_1(\mathcal{L}(\bar{Y}_t^{\gamma,n}), \mathcal{L}(\hat{Y}_t^{\gamma}))$ in (4.20), we adapt the proof from Lemma 3.28 in [9]. By Proposition 4.23, Corollary 4.20, Lemma 4.22 and 4.24, one obtains

$$\begin{aligned}
&W_1(\mathcal{L}(\bar{Y}_t^{\gamma,n}), \mathcal{L}(\hat{Y}_t^{\gamma}))\\
&\leq \sum_{k=1}^n W_1(\mathcal{L}(\bar{Y}_t^{\gamma,k}), \mathcal{L}(\bar{Y}_t^{\gamma,k-1})),\\
&\leq \sum_{k=1}^n \tilde{W}_{1,2}(\mathcal{L}(\tilde{Y}_t^{kT,\bar{\theta}_{kT}^{\gamma},\gamma}), \mathcal{L}(\tilde{Y}_t^{kT,\bar{Y}_{kT}^{\gamma,k-1},\gamma}))\\
&\leq \hat{c}\sum_{k=1}^n \exp(-\dot{c}(n-k))\tilde{W}_{1,2}(\mathcal{L}(\bar{\theta}_{kT}^{\gamma}), \mathcal{L}(\bar{Y}_{kT}^{\gamma,k-1}))\\
&\leq \hat{c}\sum_{k=1}^n \exp(-\dot{c}(n-k))W_2(\mathcal{L}(\bar{\theta}_{kT}^{\gamma}), \mathcal{L}(\bar{Y}_{kT}^{\gamma,k-1}))\left[1 + \left\{\mathbb{E}[V_4(\bar{\theta}_{kT}^{\gamma})]\right\}^{1/2} + \left\{\mathbb{E}[V_4(\bar{Y}_{kT}^{\gamma,k-1})]\right\}^{1/2}\right]\\
&\leq (\sqrt{\gamma})^{-1}\hat{c}\sum_{k=1}^n \exp(-\dot{c}(n-k))W_2^2(\mathcal{L}(\bar{\theta}_{kT}^{\gamma}), \mathcal{L}(\bar{Y}_{kT}^{\gamma,k-1}))\\
&\quad + 3\sqrt{\gamma}\hat{c}\sum_{k=1}^n \exp(-\dot{c}(n-k))\left[1 + \mathbb{E}[V_4(\bar{\theta}_{kT}^{\gamma})] + \mathbb{E}[V_4(\bar{Y}_{kT}^{\gamma,k-1})]\right]\\
&\leq \sqrt{\gamma}e^{-\min\{\dot{c},a/2\}n}n\hat{c}(e^{\min\{\dot{c},a/2\}}\bar{C}_{2,1}\mathbb{E}[V_2(\theta_0)] + 12\mathbb{E}[V_4(\theta_0)])\\
&\quad + \sqrt{\gamma}\frac{\hat{c}}{1-\exp(-\dot{c})}(\bar{C}_{2,2} + 12c_3(\gamma_{\max} + a^{-1}) + 9\mathrm{v}_4(\overline{M}_4) + 15)\\
&\leq \sqrt{\gamma}(e^{-\min\{\dot{c},a/2\}n/2}\bar{C}_{2,3}\mathbb{E}[V_4(\theta_0)] + \bar{C}_{2,4})
\end{aligned}$$

where the last inequality holds due to $e^{-\bar{\alpha}n}(n+1) \leq 1 + \bar{\alpha}^{-1}$, for $\bar{\alpha} > 0$, and we take $\bar{\alpha} = \min\{\dot{c}, a/2\}/2$, moreover,

$$\begin{aligned}
\bar{C}_{2,3} &= \hat{c}\left(1 + \frac{2}{\min\{\dot{c}, a/2\}}\right)(e^{\min\{\dot{c},a/2\}}\bar{C}_{2,1} + 12)\\
\bar{C}_{2,4} &= \frac{\hat{c}}{1-\exp(-\dot{c})}(\bar{C}_{2,2} + 12c_3(\gamma_{\max} + a^{-1}) + 9\mathrm{v}_4(\overline{M}_4) + 15)
\end{aligned} \tag{4.24}$$

with $\bar{C}_{2,1}$, $\bar{C}_{2,2}$ given in 4.23, $\hat{c}$, $\dot{c}$ given in Lemma 4.28, c_3 is given in (4.11) and $\overline{M}_4$ given in (4.17). □

By using similar arguments as in Lemma 4.25, an analogous result can be obtained in W_2 distance, which is given in the following corollary.

Corollary 4.26. *Assume **C-1**, **C-2**, **C-3** and **C-4** hold. For any* $0 < \gamma < \gamma_{\max}$ *given in* (4.5), $t \in [nT, (n+1)T]$,

$$W_2(\mathcal{L}(\bar{Y}_t^{\gamma,n}), \mathcal{L}(\hat{Y}_t^{\gamma})) \leq \gamma^{1/4}(e^{-\min\{\dot{c},a/2\}n/4}\bar{C}_{2,3}^*\mathbb{E}^{1/2}[V_4(\theta_0)] + \bar{C}_{2,4}^*),$$

where $\bar{C}_{2,3}^*$, $\bar{C}_{2,4}^*$ *is given in* (4.25).

Proof. One notices that $W_2 \leq \sqrt{2\tilde{W}_{1,2}}$, then one writes

$$W_2(\mathcal{L}(\bar{Y}_t^{\gamma,n}), \mathcal{L}(\hat{Y}_t^{\gamma}))$$

$$
\begin{aligned}
&\leq \sum_{k=1}^{n} W_2(\mathcal{L}(\bar{Y}_t^{\gamma,k}), \mathcal{L}(\bar{Y}_t^{\gamma,k-1})) \\
&\leq \sum_{k=1}^{n} \sqrt{2}\tilde{W}_{1,2}^{1/2}(\mathcal{L}(\tilde{Y}_t^{kT,\bar{\theta}_{kT}^{\gamma},\gamma}), \mathcal{L}(\tilde{Y}_t^{kT,\bar{Y}_{kT}^{\gamma,k-1},\gamma})) \\
&\leq \sqrt{2\hat{c}} \sum_{k=1}^{n} \exp(-\dot{c}(n-k)/2) W_2^{1/2}(\mathcal{L}(\bar{\theta}_{kT}^{\gamma}), \mathcal{L}(\bar{Y}_{kT}^{\gamma,k-1})) \\
&\qquad\qquad \times \left[1 + \left\{\mathbb{E}[V_4(\bar{\theta}_{kT}^{\gamma})]\right\}^{1/2} + \left\{\mathbb{E}[V_4(\bar{Y}_{kT}^{\gamma,k-1})]\right\}^{1/2}\right]^{1/2} \\
&\leq \gamma^{-1/4}\sqrt{2\hat{c}} \sum_{k=1}^{n} \exp(-\dot{c}(n-k)/2) W_2(\mathcal{L}(\bar{\theta}_{kT}^{\gamma}), \mathcal{L}(\bar{Y}_{kT}^{\gamma,k-1})) \\
&\quad + \gamma^{1/4}\sqrt{2\hat{c}} \sum_{k=1}^{n} \exp(-\dot{c}(n-k)/2) \left[1 + \left\{\mathbb{E}[V_4(\bar{\theta}_{kT}^{\gamma})]\right\}^{1/2} + \left\{\mathbb{E}[V_4(\bar{Y}_{kT}^{\gamma,k-1})]\right\}^{1/2}\right] \\
&\leq \sqrt{2\hat{c}}\gamma^{1/4} e^{-\min\{\dot{c},a/2\}n/2} n (e^{\min\{\dot{c},a/2\}/2} \bar{C}_{2,1}^{1/2} \mathbb{E}^{1/2}[V_2(\theta_0)] + 2\sqrt{2}\mathbb{E}^{1/2}[V_4(\theta_0)]) \\
&\quad + \sqrt{2\hat{c}}\gamma^{1/4} \frac{1}{1-\exp(-\dot{c}/2)} (\bar{C}_{2,2}^{1/2} + 2\sqrt{2c_3}(\gamma_{\max} + a^{-1})^{1/2} + \sqrt{3}\mathrm{v}_4^{1/2}(\overline{M}_4) + 2\sqrt{2}) \\
&\leq \gamma^{1/4}(e^{-\min\{\dot{c},a/2\}n/4} \bar{C}_{2,3}^{*} \mathbb{E}^{1/2}[V_4(\theta_0)] + \bar{C}_{2,4}^{*}),
\end{aligned}
$$

where

$$
\begin{aligned}
\bar{C}_{2,3}^{*} &= \sqrt{2\hat{c}}\left(1 + \frac{4}{\min\{\dot{c}, a/2\}}\right)(e^{\min\{\dot{c},a/2\}/2}\bar{C}_{2,1}^{1/2} + 2\sqrt{2}) \\
\bar{C}_{2,4}^{*} &= \frac{\sqrt{2\hat{c}}}{1-\exp(-\dot{c}/2)}(\bar{C}_{2,2}^{1/2} + 2\sqrt{2c_3}(\gamma_{\max} + a^{-1})^{1/2} + \sqrt{3}\mathrm{v}_4^{1/2}(\overline{M}_4) + 2\sqrt{2}),
\end{aligned}
\tag{4.25}
$$

with $\bar{C}_{2,1}$, $\bar{C}_{2,2}$ given in 4.23, $\hat{c}$, $\dot{c}$ given in Lemma 4.28, c_3 is given in (4.11) and $\overline{M}_4$ given in Lemma 4.21. This completes the proof. □

Finally, by using the inequality (4.20) and the results from previous lemmas, one can obtain the non-asymptotic bound in W_1 distance between the law of the processes $\bar{\theta}_t^{\gamma}$ and $\hat{Y}_t^{\gamma}$ for $t \in [nT, (n+1)T]$.

Lemma 4.27. *Assume **C-1**, **C-2**, **C-3** and **C-4** hold. For any $0 < \gamma < \gamma_{\max}$ given in* (4.5)*, $t \in [nT, (n+1)T]$,*

$$
W_1(\mathcal{L}(\bar{\theta}_t^{\gamma}), \mathcal{L}(\hat{Y}_t^{\gamma})) \leq \bar{C}_2\sqrt{\gamma}(e^{-\min\{\dot{c},a/2\}n/2}\mathbb{E}[V_4(\theta_0)] + 1),
$$

where $\bar{C}_2$ is given in (4.26)*.*

Proof. By using Lemma 4.24 and 4.25, one obtains

$$
\begin{aligned}
&W_1(\mathcal{L}(\bar{\theta}_t^{\gamma}), \mathcal{L}(\hat{Y}_t^{\gamma})) \\
&\leq W_1(\mathcal{L}(\bar{\theta}_t^{\gamma}), \mathcal{L}(\bar{Y}_t^{\gamma,n})) + W_1(\mathcal{L}(\bar{Y}_t^{\gamma,n}), \mathcal{L}(\hat{Y}_t^{\gamma})) \\
&\leq \sqrt{\gamma}(e^{-an/4}\bar{C}_{2,1}^{1/2}\mathbb{E}^{1/2}[V_2(\theta_0)] + \bar{C}_{2,2}^{1/2}) + \sqrt{\gamma}(e^{-\min\{\dot{c},a/2\}n/2}\bar{C}_{2,3}\mathbb{E}[V_4(\theta_0)] + \bar{C}_{2,4}) \\
&\leq \bar{C}_2\sqrt{\gamma}(e^{-\min\{\dot{c},a/2\}n/2}\mathbb{E}[V_4(\theta_0)] + 1),
\end{aligned}
$$

where

$$
\bar{C}_2 = \bar{C}_{2,1}^{1/2} + \bar{C}_{2,2}^{1/2} + \bar{C}_{2,3} + \bar{C}_{2,4}. \tag{4.26}
$$

□

Before proceeding to the proofs of the main results, we provide explicitly the constants $\dot{c}$ and $\hat{c}$ in Proposition 4.23.

Lemma 4.28. *The contraction constant in Proposition 4.23 is given by*

$$\dot{c} = \min\{\bar{\phi}, \bar{c}(p), 4\tilde{c}(p)\dot{\epsilon}\bar{c}(p)\}/2,$$

where the explicit expressions for $\bar{c}(p)$ and $\tilde{c}(p)$ can be found in Lemma 4.21 and $\bar{\phi}$ is given by

$$\bar{\phi} = \left(\sqrt{4\pi/L}\bar{b}\exp\left(\left(\bar{b}\sqrt{L}/2 + 2/\sqrt{L}\right)^2\right)\right)^{-1}.$$

Furthermore, any $\dot{\epsilon}$ can be chosen which satisfies the following inequality

$$\dot{\epsilon} \le 1 \wedge \left(8\tilde{c}(p)\sqrt{\pi/L}\int_0^{\tilde{b}} \exp\left(\left(s\sqrt{L}/2 + 2/\sqrt{L}\right)^2\right) ds\right)^{-1},$$

where $\tilde{b} = \sqrt{2\tilde{c}(p)/\bar{c}(p) - 1}$ and $\bar{b} = \sqrt{4\tilde{c}(p)(1+\bar{c}(p))/\bar{c}(p) - 1}$. The constant $\hat{c}$ is given as the ratio C_{11}/C_{10}, where C_{11}, C_{10} are given explicitly in [9, Lemma 3.26].

Proof. See [9, Lemma 3.26]. □

Proof of Theorem 4.6 One notes that, by Lemma 4.27, for $t \in [nT, (n+1)T]$

$$\begin{aligned}
W_1(\mathcal{L}(\bar{\theta}_t^\gamma), \pi_\beta) &\le W_1(\mathcal{L}(\bar{\theta}_t^\gamma), \mathcal{L}(\hat{Y}_t^\gamma)) + W_1(\mathcal{L}(\hat{Y}_t^\gamma), \pi_\beta) \\
&\le \bar{C}_2\sqrt{\gamma}(e^{-\min\{\dot{c},a/2\}n/2}\mathbb{E}[V_4(\theta_0)] + 1) + \hat{c}e^{-\dot{c}\gamma t}\tilde{W}_{1,2}(\theta_0, \pi_\beta) \\
&\le \bar{C}_2\sqrt{\gamma}(e^{-\min\{\dot{c},a/2\}n/2}\mathbb{E}[V_4(\theta_0)] + 1) \\
&\quad + \hat{c}e^{-\dot{c}\gamma t}\left[1 + \mathbb{E}[V_2(\theta_0)] + \int_{\mathbb{R}^d} V_2(\theta)\pi_\beta(d\theta)\right] \\
&\le 2e^{-\min\{\dot{c},a/2\}n/2}(\gamma_{\max}^{1/2}\bar{C}_2 + \hat{c})(1 + \mathbb{E}[|\theta_0|^4]) \\
&\quad + \hat{c}e^{-\min\{\dot{c},a/2\}n/2}\left[1 + \int_{\mathbb{R}^d} V_2(\theta)\pi_\beta(d\theta)\right] + \sqrt{\gamma}\bar{C}_2,
\end{aligned}$$

which implies, for any $n \in \mathbb{N}$

$$W_1(\mathcal{L}(\theta_n^\gamma), \pi_\beta) \le C_1 e^{-C_0\gamma n}(1 + \mathbb{E}[|\theta_0|^4]) + C_2\sqrt{\gamma},$$

where

$$\begin{aligned}
&C_0 = \dot{c}/2, \\
&C_1 = 2\left[(\gamma_{\max}^{1/2}\bar{C}_2 + \hat{c}) + \hat{c}\left(1 + \int_{\mathbb{R}^d} V_2(\theta)\pi_\beta(d\theta)\right)\right] = O\left(e^{C_* d^2/\beta^2}\left(1 + \frac{1}{1 - e^{-\dot{c}}}\right)\right), \\
&C_2 = \bar{C}_2 = O\left(e^{C_* d^2/\beta^2}\left(1 + \frac{1}{1 - e^{-\dot{c}}}\right)\right)
\end{aligned} \tag{4.27}$$

with $\bar{C}_2$ given in 4.26 and $C_* > 0$ independent of d, β, n.

Proof of Corollary 4.8 By using (4.22) in Lemma 4.24, Corollary 4.26 and Proposition 4.23, one obtains

$$\begin{aligned}
W_2(\mathcal{L}(\bar{\theta}_t^\gamma), \pi_\beta) &\le W_2(\mathcal{L}(\bar{\theta}_t^\gamma), \mathcal{L}(\hat{Y}_t^\gamma)) + W_2(\mathcal{L}(\hat{Y}_t^\gamma), \pi_\beta) \\
&\le W_2(\mathcal{L}(\bar{\theta}_t^\gamma), \mathcal{L}(\bar{Y}_t^{\gamma,n})) + W_2(\mathcal{L}(\bar{Y}_t^{\gamma,n}), \mathcal{L}(\hat{Y}_t^\gamma)) + W_2(\mathcal{L}(\hat{Y}_t^\gamma), \pi_\beta) \\
&\le \sqrt{\gamma}(e^{-an/2}\bar{C}_{2,1}\mathbb{E}[V_2(\theta_0)] + \bar{C}_{2,2})^{1/2} \\
&\quad + \gamma^{1/4}(e^{-\min\{\dot{c},a/2\}n/4}\bar{C}_{2,3}^*\mathbb{E}^{1/2}[V_4(\theta_0)] + \bar{C}_{2,4}^*), + \sqrt{2\tilde{W}_{1,2}(\mathcal{L}(\hat{Y}_t^\gamma), \pi_\beta)} \\
&\le \gamma^{1/4}\tilde{C}_2(e^{-\min\{\dot{c},a/2\}n/4}\mathbb{E}[V_4(\theta_0)] + 1) + \hat{c}^{1/2}e^{-\dot{c}\gamma t/2}\sqrt{2\tilde{W}_{1,2}(\theta_0, \pi_\beta)},
\end{aligned}$$

where $\tilde{C}_2 = \gamma_{\max}^{1/4}\bar{C}_{2,1}^{1/2} + \gamma_{\max}^{1/4}\bar{C}_{2,2}^{1/2} + \bar{C}_{2,3}^* + \bar{C}_{2,4}^*$ and it can be further calculated as

$$
\begin{aligned}
W_2(\mathcal{L}(\bar{\theta}_t^\gamma), \pi_\beta) &\leq \gamma^{1/4}\tilde{C}_2(e^{-\min\{\dot{c},a/2\}n/4}\mathbb{E}[V_4(\theta_0)] + 1) \\
&\quad + \sqrt{2}\hat{c}^{1/2}e^{-\dot{c}\gamma t/2}\left(1 + \mathbb{E}[V_2(\theta_0)] + \int_{\mathbb{R}^d} V_2(\theta)\pi_\beta(d\theta)\right)^{1/2} \\
&\leq 2e^{-\min\{\dot{c},a/2\}n/4}(\gamma_{\max}^{1/4}\tilde{C}_2 + \sqrt{2}\hat{c}^{1/2})(1 + \mathbb{E}[|\theta_0|^4]) \\
&\quad + \sqrt{2}\hat{c}^{1/2}e^{-\min\{\dot{c},a/2\}n/4}\left[1 + \int_{\mathbb{R}^d} V_2(\theta)\pi_\beta(d\theta)\right] + \gamma^{1/4}\tilde{C}_2,
\end{aligned}
$$

Finally, one obtains

$$
W_2(\mathcal{L}(\theta_n^\gamma), \pi_\beta) \leq C_4 e^{-C_3\gamma n}\mathbb{E}[|\theta_0|^4 + 1] + C_5\gamma^{1/4}
$$

where

$$
\begin{aligned}
C_3 &= \dot{c}/4, \\
C_4 &= 2\left[(\gamma_{\max}^{1/4}\tilde{C}_2 + \sqrt{2}\hat{c}^{1/2}) + \hat{c}^{1/2}\left(1 + \int_{\mathbb{R}^d} V_2(\theta)\pi_\beta(d\theta)\right)\right] \\
&= O\left(e^{C_* d^2/\beta^2}\left(1 + \frac{1}{1 - e^{-\dot{c}/2}}\right)\right), \\
C_5 &= \tilde{C}_2 = O\left(e^{C_* d^2/\beta^2}\left(1 + \frac{1}{1 - e^{-\dot{c}/2}}\right)\right)
\end{aligned}
\tag{4.28}
$$

with $C_* > 0$ independent of d, β, n.
Proof of Corollary 4.9 To obtain an upper bound for the expected excess risk $\mathbb{E}[\hat{U}(\hat{\theta})] - \inf_{\theta\in\mathbb{R}^d}\hat{U}(\theta)$, one considers the following splitting

$$
\mathbb{E}[\hat{U}(\hat{\theta})] - \inf_{\theta\in\mathbb{R}^d}\hat{U}(\theta) = \left(\mathbb{E}[\hat{U}(\hat{\theta})] - \mathbb{E}[\hat{U}(\hat{Y}_\infty)]\right) + \left(\mathbb{E}[\hat{U}(\hat{Y}_\infty)] - \inf_{\theta\in\mathbb{R}^d}\hat{U}(\theta)\right), \tag{4.29}
$$

where $\hat{\theta} = \theta_n^\gamma$ and $\hat{Y}_\infty \sim \pi_\beta$ with $\pi_\beta(\theta) = \exp(-\beta\hat{U}(\theta))$ for all $\theta \in \mathbb{R}^d$. By using [42, Lemma 3.5], Lemma 4.18, C.3 and Corollary 4.8, the first term on the RHS of (4.29) can be bounded by

$$
\begin{aligned}
&\mathbb{E}[\hat{U}(\hat{\theta})] - \mathbb{E}[\hat{U}(\hat{Y}_\infty)] \\
&\leq \left(L(\mathbb{E}\left[|\theta_0|^2\right] + (c_1 + \mathbb{E}[K_1^2(Z_0)]/a)(\gamma_{\max} + a^{-1}))^{1/2} + |h(0)|\right) W_2(\mathcal{L}(\theta_n^\gamma), \pi_\beta) \\
&\leq \left(L(\mathbb{E}\left[|\theta_0|^2\right] + (c_1 + \mathbb{E}[K_1^2(Z_0)]/a)(\gamma_{\max} + a^{-1}))^{1/2} + |h(0)|\right)\left(C_4 e^{-C_3\gamma n}\mathbb{E}[|\theta_0|^4 + 1] + C_5\gamma^{1/4}\right) \\
&\leq \hat{C}_1 e^{-\hat{C}_0\gamma n} + \hat{C}_2\gamma^{1/4},
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{C}_0 &= C_3, \\
\hat{C}_1 &= C_4\left(L(\mathbb{E}\left[|\theta_0|^2\right] + (c_1 + \mathbb{E}[K_1^2(Z_0)]/a)(\gamma_{\max} + a^{-1}))^{1/2} + |h(0)|\right)\mathbb{E}[|\theta_0|^4 + 1] \\
&= O\left(e^{C_* d^2/\beta^2}\left(1 + \frac{1}{1 - e^{-\dot{c}/2}}\right)\right), \\
\hat{C}_2 &= C_5\left(L(\mathbb{E}\left[|\theta_0|^2\right] + (c_1 + \mathbb{E}[K_1^2(Z_0)]/a)(\gamma_{\max} + a^{-1}))^{1/2} + |h(0)|\right) \\
&= O\left(e^{C_* d^2/\beta^2}\left(1 + \frac{1}{1 - e^{-\dot{c}/2}}\right)\right)
\end{aligned}
\tag{4.30}
$$

with C_3, C_4, C_5 given in (4.28), c_1 given in (4.10) and $C_* > 0$ independent of d, β, n. Moreover, the second term on the RHS of (4.29) can be estimated by using [42, Proposition 3.4], which gives,

$$\mathbb{E}[\hat{U}(\hat{Y}_\infty)] - \inf_{\theta\in\mathbb{R}^d} \hat{U}(\theta) \le \frac{\hat{C}_3}{\beta},$$

where

$$\hat{C}_3 = \frac{d}{2}\log\left(\frac{e\beta L}{ad}\left(\frac{2d}{\beta} + 2b + \frac{\mathbb{E}[K_1^2(Z_0)]}{a}\right)\right) = O\left(d\log\left(C_*\left(\frac{\beta}{d}+1\right)\right)\right). \tag{4.31}$$

Finally, one obtains

$$\mathbb{E}[\hat{U}(\hat{\theta})] - \inf_{\theta\in\mathbb{R}^d} \hat{U}(\theta) \le \hat{C}_1 e^{-\hat{C}_0\gamma n} + \hat{C}_2\gamma^{1/4} + \hat{C}_3/\beta.$$

4.4 Proof of the main results: convex case

To obtain the convergence result of the SGLD algorithm (4.1) in the convex case, i.e. Theorem 4.15, we first introduce the LMC algorithm associated with SDE (4.2), which is given explicitly by, for any $n \in \mathbb{N}$,

$$\dot{\theta}^\gamma_{n+1} := \dot{\theta}^\gamma_n - \gamma h(\dot{\theta}^\gamma_n) + \sqrt{2\beta^{-1}\gamma}\xi_{n+1}, \quad \dot{\theta}^\gamma_0 := \theta_0. \tag{4.32}$$

For $0 < \gamma < \bar{\gamma}_{\max}$ given in (4.8), the Markov kernel $\dot{R}_\gamma$ associated with (4.32) is given by, for all $A \in \mathscr{B}(\mathbb{R}^d)$ and $\theta \in \mathbb{R}^d$,

$$\dot{R}_\gamma(\theta, A) = \int_A (4\beta^{-1}\pi\gamma)^{-d/2}\exp\left(-\beta(4\gamma)^{-1}\left|y-\theta+\gamma h(\theta)\right|^2\right) dy.$$

In this section, the moment estimates of SDE (4.2), the LMC algorithm (4.32) and the SGLD algorithm (4.1) are presented which are then used in the analysis of the convergence results.

4.4.1 Preliminary estimates

By **C-5** and **C-6**, $\hat{U}$ has a unique minimizer $\theta^* \in \mathbb{R}^d$. Denote by $(P_t)_{t\ge0}$ the semigroup associated with SDE (4.2). The statements below provide a moment estimate and a convergence result of SDE (4.2).

Lemma 4.29 (Proposition 1 in [15]). *Assume **C-1**, **C-2**, **C-3**, **C-5** and **C-6** hold.*

(i) For all $t > 0$ and $y \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} |y-\theta^*|^2 P_t(\theta, dy) \le |\theta-\theta^*|^2 e^{-2\hat{a}t} + (d/(\hat{a}\beta))(1-e^{-2\hat{a}t}).$$

(ii) The stationary distribution π_β satisfies

$$\int_{\mathbb{R}^d} |y-\theta^*|^2 \pi_\beta(dy) \le d/(\hat{a}\beta).$$

The following lemma provides moment estimates for $(\dot{\theta}_n)_{n\in\mathbb{N}}$ and it states that $\dot{R}_\gamma$ admits an invariant measure π_γ which may differ from π_β.

Lemma 4.30 (Proposition 2 and Proposition 3 (ii) in [15]). *Assume **C-1**, **C-2**, **C-5** and **C-6** hold. Then, for all $0 < \gamma < \bar{\gamma}_{\max}$ given in (4.8), one obtains:*

(i) For all $t > 0$ and $\theta \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} |y-\theta^*|^2 \dot{R}^n_\gamma(\theta, dy) \le (1-2\hat{a}^*\gamma)^n|\theta-\theta^*|^2 + (d/(\hat{a}^*\beta))(1-(1-2\hat{a}^*\gamma)^n),$$

where $\hat{a}^* = \hat{a}L/(\hat{a}+L)$.

(ii) The Markov kernel $\dot{R}_\gamma$ has a unique stationary distribution π_γ and it satisfies

$$\int_{\mathbb{R}^d} |\theta - \theta^*|^2 \pi_\gamma(d\theta) \leq d/(\hat{a}^*\beta).$$

(iii) For all $n \in \mathbb{R}^d$ and $\theta \in \mathbb{R}^d$,

$$W_2(\delta_\theta \dot{R}_\gamma^n, \pi_\gamma) \leq e^{-\hat{a}^*\gamma n}(|\theta - \theta^*|^2 + d/(\hat{a}^*\beta))^{1/2}.$$

The lemma below presents a second moment bound for θ_n^γ in the convex case.

Lemma 4.31. *Assume **C-1**, **C-2**, **C-5** hold. For any $0 < \gamma < \bar{\gamma}_{\max}$ given in* (4.8)*,*

$$\mathbb{E}\left[|\theta_n^\gamma - \theta^*|^2\right] \leq (1 - \hat{a}\gamma)^n \mathbb{E}\left[|\theta_0^\gamma - \theta^*|^2\right] + \bar{c}_4 \hat{a}^{-1},$$

where

$$\bar{c}_4 = 32\mathbb{E}\left[K_1^2(Z_0)\right]\hat{a}_1^{-1} + 9\bar{\gamma}_{\max}(L_1^2\mathbb{E}\left[K_\rho(Z_0)\right]|\theta^*|^2 + L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^2(Z_0)\right]) + 2d\beta^{-1}. \tag{4.33}$$

*This implies $\sup_n \mathbb{E}\left[\left|\theta_{n+1}^\gamma - \theta^*\right|^2\right] \leq \mathbb{E}\left[|\theta_0^\gamma - \theta^*|^2\right] + \bar{c}_4\hat{a}^{-1} < \infty$. Furthermore, if $\rho = 0$ in **C-1**, the result holds for $\gamma \in \min\{1/2(\hat{a} + L), 1/(6L_1)\}$ with $\hat{a} = \hat{a}_1 + \hat{a}_2$.*

Proof. By using (4.1), one writes, for any $n \in \mathbb{N}$,

$$\begin{aligned}
|\theta_{n+1}^\gamma - \theta^*|^2 &= |\theta_n^\gamma - \theta^*|^2 + 2\left\langle \theta_n^\gamma - \theta^*, -\gamma H(\theta_n^\gamma, Z_{n+1}) + \sqrt{2\beta^{-1}\gamma}\xi_{n+1}\right\rangle \\
&\quad + |-\gamma H(\theta_n^\gamma, Z_{n+1}) + \sqrt{2\beta^{-1}\gamma}\xi_{n+1}|^2 \\
&= |\theta_n^\gamma - \theta^*|^2 - 2\gamma\left\langle \theta_n^\gamma - \theta^*, H(\theta_n^\gamma, Z_{n+1}) - H(\theta^*, Z_{n+1})\right\rangle \\
&\quad - 2\gamma\left\langle \theta_n^\gamma - \theta^*, H(\theta^*, Z_{n+1})\right\rangle + 2\left\langle \theta_n^\gamma - \theta^*, \sqrt{2\beta^{-1}\gamma}\xi_{n+1}\right\rangle \\
&\quad + \gamma^2|H(\theta_n^\gamma, Z_{n+1})|^2 - 2\gamma\left\langle H(\theta_n^\gamma, Z_{n+1}), \sqrt{2\beta^{-1}\gamma}\xi_{n+1}\right\rangle + 2\beta^{-1}\gamma|\xi_{n+1}|^2.
\end{aligned}$$

Taking conditional expectation on both sides, and by using Remark 4.1, **C-1** and **C-5** yield

$$\begin{aligned}
&\mathbb{E}\left[|\theta_{n+1}^\gamma - \theta^*|^2 \middle| \theta_n^\gamma\right] \\
&= |\theta_n^\gamma - \theta^*|^2 - 2\gamma\mathbb{E}\left[\left\langle \theta_n^\gamma - \theta^*, F(\theta_n^\gamma, Z_{n+1}) - F(\theta^*, Z_{n+1})\right\rangle \middle| \theta_n^\gamma\right] \\
&\quad - 2\gamma\mathbb{E}\left[\left\langle \theta_n^\gamma - \theta^*, G(\theta_n^\gamma, Z_{n+1}) - G(\theta^*, Z_{n+1})\right\rangle \middle| \theta_n^\gamma\right] \\
&\quad - 2\gamma\left\langle \theta_n^\gamma - \theta^*, h(\theta^*)\right\rangle + \gamma^2\mathbb{E}\left[|H(\theta_n^\gamma, Z_{n+1})|^2 \middle| \theta_n^\gamma\right] + 2d\beta^{-1}\gamma \\
&\leq |\theta_n^\gamma - \theta^*|^2 - 2\gamma\hat{a}_1|\theta_n^\gamma - \theta^*|^2 + 4\gamma\mathbb{E}\left[K_1(Z_0)\right]|\theta_n^\gamma - \theta^*| \\
&\quad + \gamma^2\mathbb{E}\left[\left((1 + |Z_{n+1}|)^{\rho+1}(L_1|\theta_n^\gamma - \theta^*| + L_1|\theta^*| + L_2) + F_*(Z_{n+1})\right)^2 \middle| \theta_n^\gamma\right] + 2d\beta^{-1}\gamma \\
&\leq (1 - 2\hat{a}_1\gamma)|\theta_n^\gamma - \theta^*|^2 + 4\gamma\mathbb{E}\left[K_1(Z_0)\right]|\theta_n^\gamma - \theta^*| + 2\gamma^2L_1^2\mathbb{E}\left[K_\rho(Z_0)\right]|\theta_n^\gamma - \theta^*|^2 \\
&\quad + 6\gamma^2L_1^2\mathbb{E}\left[K_\rho(Z_0)\right]|\theta^*|^2 + 6\gamma^2L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 6\gamma^2\mathbb{E}\left[F_*^2(Z_0)\right] + 2d\beta^{-1}\gamma
\end{aligned} \tag{4.34}$$

which implies, for $0 < \gamma < \bar{\gamma}_{\max}$,

$$\begin{aligned}
&\mathbb{E}\left[|\theta_{n+1}^\gamma - \theta^*|^2 \middle| \theta_n^\gamma\right] \\
&\leq \left(1 - \frac{3}{2}\hat{a}_1\gamma\right)|\theta_n^\gamma - \theta^*|^2 + 4\gamma\mathbb{E}\left[K_1(Z_0)\right]|\theta_n^\gamma - \theta^*| \\
&\quad + 6\gamma^2L_1^2\mathbb{E}\left[K_\rho(Z_0)\right]|\theta^*|^2 + 6\gamma^2L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 6\gamma^2\mathbb{E}\left[F_*^2(Z_0)\right] + 2d\beta^{-1}\gamma.
\end{aligned}$$

Then, for $|\theta_n^\gamma - \theta^*| > 8\mathbb{E}\left[K_1(Z_0)\right]\hat{a}_1^{-1}$, one notices that

$$-\frac{1}{2}\hat{a}_1\gamma|\theta_n^\gamma - \theta^*|^2 + 4\gamma\mathbb{E}\left[K_1(Z_0)\right]|\theta_n^\gamma - \theta^*| < 0,$$

and this indicates

$$\mathbb{E}\left[|\theta_{n+1}^\gamma - \theta^*|^2 \middle| \theta_n^\gamma\right] \leq (1 - \hat{a}_1\gamma)|\theta_n^\gamma - \theta^*|^2 + 6\gamma^2 L_1^2 \mathbb{E}\left[K_\rho(Z_0)\right]|\theta^*|^2 \\ + 6\gamma^2 L_2^2 \mathbb{E}\left[K_\rho(Z_0)\right] + 6\gamma^2 \mathbb{E}\left[F_*^2(Z_0)\right] + 2d\beta^{-1}\gamma.$$

Similarly, for $|\theta_n^\gamma - \theta^*| \leq 8\mathbb{E}\left[K_1(Z_0)\right]\hat{a}_1^{-1}$, one obtains

$$\begin{aligned} &\mathbb{E}\left[|\theta_{n+1}^\gamma - \theta^*|^2 \middle| \theta_n^\gamma\right] \\ &\leq \left(1 - \frac{3}{2}\hat{a}_1\gamma\right)|\theta_n^\gamma - \theta^*|^2 + 32\gamma\mathbb{E}\left[K_1^2(Z_0)\right]\hat{a}_1^{-1} \\ &\quad + 6\gamma^2 L_1^2 \mathbb{E}\left[K_\rho(Z_0)\right]|\theta^*|^2 + 6\gamma^2 L_2^2 \mathbb{E}\left[K_\rho(Z_0)\right] + 6\gamma^2 \mathbb{E}\left[F_*^2(Z_0)\right] + 2d\beta^{-1}\gamma. \end{aligned}$$

Combining the two cases yields

$$\mathbb{E}\left[|\theta_{n+1}^\gamma - \theta^*|^2 \middle| \theta_n^\gamma\right] \leq (1 - \hat{a}\gamma)|\theta_n^\gamma - \theta^*|^2 + \gamma c_4,$$

where $c_4 = 32\mathbb{E}\left[K_1^2(Z_0)\right]\hat{a}_1^{-1} + 6\bar{\gamma}_{\max}(L_1^2\mathbb{E}\left[K_\rho(Z_0)\right]|\theta^*|^2 + L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^2(Z_0)\right]) + 2d\beta^{-1}$. The result follows by induction.

Moreover, one observes that when $\rho = 0$ in **C-1**, F is co-coercive, i.e. for any $\theta, \theta' \in \mathbb{R}^d$ and for every $z \in \mathbb{R}^m$

$$\langle \theta - \theta', F(\theta, z) - F(\theta', z)\rangle \geq \frac{1}{L_1}|F(\theta, z) - F(\theta', z)|^2. \tag{4.35}$$

Then, by substituting (4.35) into (4.34), one obtains

$$\begin{aligned} \mathbb{E}\left[|\theta_{n+1}^\gamma - \theta^*|^2 \middle| \theta_n^\gamma\right] &\leq |\theta_n^\gamma - \theta^*|^2 - \frac{3}{2}\gamma\hat{a}_1|\theta_n^\gamma - \theta^*|^2 - \frac{\gamma}{2L_1}\mathbb{E}\left[|F(\theta_n^\gamma, Z_{n+1}) - F(\theta^*, Z_{n+1})|^2 \middle| \theta_n^\gamma\right] \\ &\quad + 4\gamma\mathbb{E}\left[K_1(Z_0)\right]|\theta_n^\gamma - \theta^*| + \gamma^2\mathbb{E}\left[|H(\theta_n^\gamma, Z_{n+1})|^2 \middle| \theta_n^\gamma\right] + 2d\beta^{-1}\gamma \\ &\leq \left(1 - \frac{3}{2}\gamma\hat{a}_1\right)|\theta_n^\gamma - \theta^*|^2 + 4\gamma\mathbb{E}\left[K_1(Z_0)\right]|\theta_n^\gamma - \theta^*| \\ &\quad + \left(3\gamma^2 - \frac{\gamma}{2L_1}\right)\mathbb{E}\left[|F(\theta_n^\gamma, Z_{n+1}) - F(\theta^*, Z_{n+1})|^2 \middle| \theta_n^\gamma\right] \\ &\quad + 3\gamma^2\mathbb{E}\left[|F(\theta^*, Z_{n+1})|^2 \middle| \theta_n^\gamma\right] + 3\gamma^2\mathbb{E}\left[K_1^2(Z_0)\right] + 2d\beta^{-1}\gamma, \end{aligned}$$

which implies for $\gamma \in \min\{1/2\hat{a}_1, 1/(6L_1)\}$

$$\begin{aligned} &\mathbb{E}\left[|\theta_{n+1}^\gamma - \theta^*|^2 \middle| \theta_n^\gamma\right] \\ &\leq \left(1 - \frac{3}{2}\gamma\hat{a}_1\right)|\theta_n^\gamma - \theta^*|^2 + 4\gamma\mathbb{E}\left[K_1(Z_0)\right]|\theta_n^\gamma - \theta^*| \\ &\quad + 9\gamma^2 L_1^2\mathbb{E}\left[K_\rho(Z_0)\right]|\theta^*|^2 + 9\gamma^2 L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 9\gamma^2\mathbb{E}\left[F_*^2(Z_0)\right] + 2d\beta^{-1}\gamma. \end{aligned}$$

By using the same arguments as above, consider the case $|\theta_n^\gamma - \theta^*| > 8\mathbb{E}\left[K_1(Z_0)\right]\hat{a}_1^{-1}$, one notices that

$$-\frac{1}{2}\hat{a}_1\gamma|\theta_n^\gamma - \theta^*|^2 + 4\gamma\mathbb{E}\left[K_1(Z_0)\right]|\theta_n^\gamma - \theta^*| < 0,$$

and this indicates

$$\mathbb{E}\left[|\theta_{n+1}^\gamma - \theta^*|^2 \middle| \theta_n^\gamma\right] \leq (1 - \hat{a}_1\gamma)|\theta_n^\gamma - \theta^*|^2 + 9\gamma^2 L_1^2\mathbb{E}\left[K_\rho(Z_0)\right]|\theta^*|^2 \\ + 9\gamma^2 L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 9\gamma^2\mathbb{E}\left[F_*^2(Z_0)\right] + 2d\beta^{-1}\gamma.$$

Similarly, for $|\theta_n^\gamma - \theta^*| \leq 8\mathbb{E}\left[K_1(Z_0)\right]\hat{a}_1^{-1}$, one obtains

$$\mathbb{E}\left[|\theta_{n+1}^\gamma - \theta^*|^2 \middle| \theta_n^\gamma\right] \leq \left(1 - \frac{3}{2}\hat{a}_1\gamma\right)|\theta_n^\gamma - \theta^*|^2 + 32\gamma\mathbb{E}\left[K_1^2(Z_0)\right]\hat{a}_1^{-1}$$

$$+ 9\gamma^2 L_1^2 \mathbb{E}\left[K_\rho(Z_0)\right]|\theta^*|^2 + 9\gamma^2 L_2^2 \mathbb{E}\left[K_\rho(Z_0)\right] + 9\gamma^2 \mathbb{E}\left[F_*^2(Z_0)\right] + 2d\beta^{-1}\gamma.$$

Combining the two cases yields

$$\mathbb{E}\left[|\theta_{n+1}^\gamma - \theta^*|^2 \middle| \theta_n^\gamma\right] \leq (1 - \hat{a}\gamma)|\theta_n^\gamma - \theta^*|^2 + \gamma \bar{c}_4,$$

where $\bar{c}_4 = 32\mathbb{E}\left[K_1^2(Z_0)\right]\hat{a}_1^{-1} + 9\bar{\gamma}_{\max}(L_1^2\mathbb{E}\left[K_\rho(Z_0)\right]|\theta^*|^2 + L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + \mathbb{E}\left[F_*^2(Z_0)\right]) + 2d\beta^{-1}$. $\square$

4.4.2 Convergence results

We aim to establish the non-asymptotic bound in Wasserstein-2 distance between $\mathcal{L}(\theta_n^\gamma)$ and π_β. To achieve this, we consider the following decomposition:

$$W_2(\mathcal{L}(\theta_n^\gamma), \pi_\beta) \leq W_2(\mathcal{L}(\theta_n^\gamma), \mathcal{L}(\dot{\theta}_n^\gamma)) + W_2(\mathcal{L}(\dot{\theta}_n^\gamma), \pi_\gamma) + W_2(\pi_\gamma, \pi_\beta). \tag{4.36}$$

The lemma presented below provides the non-asymptotic estimates for the last two terms in (4.36).

Theorem 4.32. *[15, Corollary 7] Assume **C-1**, **C-2**, **C-3**, **C-5** and **C-6** hold. Then, for any $0 < \gamma < \bar{\gamma}_{\max}$ given in (4.8), the Markov chain $(\dot{\theta}_n^\gamma)_{n\in\mathbb{N}}$ admits an invariant measure π_γ such that, for all $n \in \mathbb{N}$,*

$$W_2(\mathcal{L}(\dot{\theta}_n^\gamma), \pi_\gamma) \leq \bar{C}_7 e^{-\hat{a}^*\gamma n},$$

where $\bar{C}_7 = (|\theta_0 - \theta|^2 + d/(\hat{a}^*\beta))^{1/2}$ is given in Lemma 4.30 (iii) with $\hat{a}^* = \hat{a}L/(\hat{a} + L)$. Furthermore,

$$W_2(\pi_\beta, \pi_\gamma) \leq \bar{C}_{8,1}\sqrt{\gamma},$$

where

$$\bar{C}_{8,1} = \left(dL^2(\hat{a}^*\beta)^{-1}(2\gamma + (\hat{a}^*)^{-1})(1 + \tfrac{1}{12}\gamma^2 L^2 + \tfrac{1}{2}L^2\gamma/\hat{a})\right)^{1/2}. \tag{4.37}$$

The non-asymptotic estimate for the first term in (4.36) is provided in the following lemma.

Lemma 4.33. *Assume **C-1**, **C-2**, **C-3**, **C-5** and **C-6** hold. For any $0 < \gamma < \bar{\gamma}_{\max}$ given in (4.8), one obtains*

$$W_2(\mathcal{L}(\dot{\theta}_n^\gamma), \mathcal{L}(\theta_n^\gamma)) \leq \bar{C}_{8,2}\sqrt{\gamma},$$

where

$$\begin{aligned} \bar{C}_{8,2} &= \sqrt{c_5/2\hat{a}^*} \\ c_5 &= (8L^2 + 16L_1^2\mathbb{E}\left[K_\rho(Z_0)\right])(\mathbb{E}\left[|\theta_0^\gamma|^2\right] + \hat{a}^{-1}\bar{c}_4) \\ &\quad + (8L^2 + 40L_1^2\mathbb{E}\left[K_\rho(Z_0)\right])|\theta^*|^2 + 24L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 24\mathbb{E}\left[F_*^2(Z_0)\right]. \end{aligned} \tag{4.38}$$

Proof. By using synchronous coupling for the algorithms (4.32) and (4.1), one obtains

$$\begin{aligned} |\dot{\theta}_{n+1}^\gamma - \theta_{n+1}^\gamma|^2 &= |\dot{\theta}_n^\gamma - \theta_n^\gamma - \gamma(h(\dot{\theta}_n^\gamma) - H(\theta_n^\gamma, Z_{n+1}))|^2 \\ &= |\dot{\theta}_n^\gamma - \theta_n^\gamma|^2 - 2\gamma\langle \dot{\theta}_n^\gamma - \theta_n^\gamma, h(\dot{\theta}_n^\gamma) - H(\theta_n^\gamma, Z_{n+1})\rangle + \gamma^2|h(\dot{\theta}_n^\gamma) - H(\theta_n^\gamma, Z_{n+1})|^2 \\ &\leq |\dot{\theta}_n^\gamma - \theta_n^\gamma|^2 - 2\gamma\langle \dot{\theta}_n^\gamma - \theta_n^\gamma, h(\dot{\theta}_n^\gamma) - h(\theta_n^\gamma)\rangle - 2\gamma\langle \dot{\theta}_n^\gamma - \theta_n^\gamma, h(\theta_n^\gamma) - H(\theta_n^\gamma, Z_{n+1})\rangle \\ &\quad + 2\gamma^2|h(\dot{\theta}_n^\gamma) - h(\theta_n^\gamma)|^2 + 2\gamma^2|h(\theta_n^\gamma) - H(\theta_n^\gamma, Z_{n+1})|^2, \end{aligned}$$

which implies, by taking conditional expectation on both sides and by using Remark 4.13

$$\begin{aligned} \mathbb{E}\left[|\dot{\theta}_{n+1}^\gamma - \theta_{n+1}^\gamma|^2 \middle| \dot{\theta}_n^\gamma, \theta_n^\gamma\right] &\leq |\dot{\theta}_n^\gamma - \theta_n^\gamma|^2 - 2\hat{a}^*\gamma|\dot{\theta}_n^\gamma - \theta_n^\gamma|^2 - \frac{2\gamma}{\hat{a} + L}|h(\dot{\theta}_n^\gamma) - h(\theta_n^\gamma)|^2 \\ &\quad + 2\gamma^2|h(\dot{\theta}_n^\gamma) - h(\theta_n^\gamma)|^2 + 2\gamma^2\mathbb{E}\left[|h(\theta_n^\gamma) - H(\theta_n^\gamma, Z_{n+1})|^2 \middle| \dot{\theta}_n^\gamma, \theta_n^\gamma\right], \end{aligned}$$

where $\hat{a}^* = \hat{a}L/(\hat{a}+L)$. For $\gamma < \bar{\gamma}_{\max}$, one obtains by using Remark 4.1 and 4.2

$$\begin{aligned}
&\mathbb{E}\left[|\dot{\theta}^\gamma_{n+1} - \theta^\gamma_{n+1}|^2\Big|\,\dot{\theta}^\gamma_n, \theta^\gamma_n\right] \\
&\leq (1-2\hat{a}^*\gamma)|\dot{\theta}^\gamma_n - \theta^\gamma_n|^2 + 4\gamma^2\mathbb{E}\left[|h(\theta^\gamma_n)|^2\Big|\,\dot{\theta}^\gamma_n, \theta^\gamma_n\right] \\
&\quad + 4\gamma^2\mathbb{E}\left[|H(\theta^\gamma_n, Z_{n+1})|^2\Big|\,\dot{\theta}^\gamma_n, \theta^\gamma_n\right] \\
&\leq (1-2\hat{a}^*\gamma)|\dot{\theta}^\gamma_n - \theta^\gamma_n|^2 + 4\gamma^2L^2\mathbb{E}\left[|\theta^\gamma_n - \theta^*|^2\Big|\,\dot{\theta}^\gamma_n, \theta^\gamma_n\right] \\
&\quad + 4\gamma^2\mathbb{E}\left[\left((1+|Z_{n+1}|)^{\rho+1}(L_1|\theta^\gamma_n - \theta^*| + L_1|\theta^*| + L_2) + F_*(Z_{n+1})\right)^2\Big|\,\dot{\theta}^\gamma_n, \theta^\gamma_n\right] \\
&\leq (1-2\hat{a}^*\gamma)|\dot{\theta}^\gamma_n - \theta^\gamma_n|^2 + (4\gamma^2L^2 + 8\gamma^2L_1^2\mathbb{E}\left[K_\rho(Z_0)\right])|\theta^\gamma_n - \theta^*|^2 \\
&\quad + 24\gamma^2L_1^2\mathbb{E}\left[K_\rho(Z_0)\right]|\theta^*|^2 + 24\gamma^2L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 24\gamma^2\mathbb{E}\left[F_*^2(Z_0)\right].
\end{aligned}$$

Finally, one calculates by taking expectations on both sides and by using Lemma 4.31,

$$\begin{aligned}
&\mathbb{E}\left[|\dot{\theta}^\gamma_{n+1} - \theta^\gamma_{n+1}|^2\right] \\
&\leq (1-2\hat{a}^*\gamma)\mathbb{E}\left[|\dot{\theta}^\gamma_n - \theta^\gamma_n|^2\right] + (4\gamma^2L^2 + 8\gamma^2L_1^2\mathbb{E}\left[K_\rho(Z_0)\right])\mathbb{E}\left[|\theta^\gamma_n - \theta^*|^2\right] \\
&\quad + 24\gamma^2L_1^2\mathbb{E}\left[K_\rho(Z_0)\right]|\theta^*|^2 + 24\gamma^2L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 24\gamma^2\mathbb{E}\left[F_*^2(Z_0)\right] \\
&\leq (1-2\hat{a}^*\gamma)\mathbb{E}\left[|\dot{\theta}^\gamma_n - \theta^\gamma_n|^2\right] + \gamma^2 c_5,
\end{aligned}$$

where

$$\begin{aligned}
c_5 &= (8L^2 + 16L_1^2\mathbb{E}\left[K_\rho(Z_0)\right])(\mathbb{E}\left[|\theta^\gamma_0|^2\right] + \hat{a}^{-1}\bar{c}_4) + (8L^2 + 40L_1^2\mathbb{E}\left[K_\rho(Z_0)\right])|\theta^*|^2 \\
&\quad + 24L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 24\mathbb{E}\left[F_*^2(Z_0)\right].
\end{aligned}$$

The result follows by induction. □

Proof of Theorem 4.15 One observes that by using Theorem 4.32 and Lemma 4.33

$$\begin{aligned}
W_2(\mathcal{L}(\theta^\gamma_n), \pi_\beta) &\leq W_2(\mathcal{L}(\theta^\gamma_n), \mathcal{L}(\dot{\theta}^\gamma_n)) + W_2(\mathcal{L}(\dot{\theta}^\gamma_n), \pi_\gamma) + W_2(\pi_\gamma, \pi_\beta) \\
&\leq \bar{C}_{8,2}\sqrt{\gamma} + \bar{C}_7 e^{-\hat{a}^*\gamma n} + \bar{C}_{8,1}\sqrt{\gamma} \\
&\leq C_7 e^{-C_6\gamma n} + C_8\sqrt{\gamma},
\end{aligned}$$

where

$$C_6 = \hat{a}^*, \quad C_7 = \bar{C}_7 = O\left(\sqrt{1+\frac{d}{\beta}}\right), \quad C_8 = \bar{C}_{8,1} + \bar{C}_{8,2} = O\left(\sqrt{1+\frac{d}{\beta}}\right) \tag{4.39}$$

with $\hat{a}^* = \hat{a}L/(\hat{a}+L)$, $\bar{C}_7$ given in Lemma 4.32, $\bar{C}_{8,1}$ and $\bar{C}_{8,2}$ given in (4.37) and (4.38) respectively.

Proof of Corollary 4.16 The proof follows the same lines as the proof of Corollary 4.9. To obtain an upper bound for the expected excess risk $\mathbb{E}[\hat{U}(\hat{\theta})] - \inf_{\theta\in\mathbb{R}^d}\hat{U}(\theta)$, one considers

$$\mathbb{E}[\hat{U}(\hat{\theta})] - \inf_{\theta\in\mathbb{R}^d}\hat{U}(\theta) = \left(\mathbb{E}[\hat{U}(\hat{\theta})] - \mathbb{E}[\hat{U}(\hat{Y}_\infty)]\right) + \left(\mathbb{E}[\hat{U}(\hat{Y}_\infty)] - \inf_{\theta\in\mathbb{R}^d}\hat{U}(\theta)\right), \tag{4.40}$$

where $\hat{\theta} = \theta^\gamma_n$ and $\hat{Y}_\infty \sim \pi_\beta$ with $\pi_\beta(\theta) = \exp(-\beta\hat{U}(\theta))$ for all $\theta \in \mathbb{R}^d$. By using [42, Lemma 3.5], Lemma 4.29, 4.31 and Theorem 4.15, the first term on the RHS of (4.40) can be bounded by

$$\begin{aligned}
&\mathbb{E}[\hat{U}(\hat{\theta})] - \mathbb{E}[\hat{U}(\hat{Y}_\infty)] \\
&\leq \left(L(\mathbb{E}\left[|\theta_0 - \theta^*|^2\right] + \bar{c}_4\hat{a}^{-1} + |\theta^*|^2)^{1/2} + |h(0)|\right) W_2(\mathcal{L}(\theta^\gamma_n), \pi_\beta) \\
&\leq \left(L(\mathbb{E}\left[|\theta_0 - \theta^*|^2\right] + \bar{c}_4\hat{a}^{-1} + |\theta^*|^2)^{1/2} + |h(0)|\right)\left(C_7 e^{-C_6\gamma n} + C_8\sqrt{\gamma}\right)
\end{aligned}$$

$$\leq \hat{C}_5 e^{-\hat{C}_4 \gamma n} + \hat{C}_6 \sqrt{\gamma},$$

where $\theta^* \in \mathbb{R}^d$ is the minimizer of $\hat{U}$, and

$$\begin{aligned}
\hat{C}_4 &= C_6, \\
\hat{C}_5 &= C_7 \left(L(\mathbb{E}\left[|\theta_0 - \theta^*|^2\right] + \bar{c}_4 \hat{a}^{-1} + |\theta^*|^2)^{1/2} + |h(0)| \right) = O\left(1 + \frac{d}{\beta}\right), \\
\hat{C}_6 &= C_8 \left(L(\mathbb{E}\left[|\theta_0 - \theta^*|^2\right] + \bar{c}_4 \hat{a}^{-1} + |\theta^*|^2)^{1/2} + |h(0)| \right) = O\left(1 + \frac{d}{\beta}\right),
\end{aligned} \tag{4.41}$$

with C_6, C_7, C_8 given in (4.39) and $\bar{c}_4$ given in (4.33). Moreover, the second term on the RHS of (4.40) can be estimated by using [42, Proposition 3.4], which gives,

$$\mathbb{E}[\hat{U}(\hat{Y}_\infty)] - \inf_{\theta \in \mathbb{R}^d} \hat{U}(\theta) \leq \frac{\hat{C}_7}{\beta},$$

where

$$\hat{C}_7 = \frac{d}{2} \log \left(\frac{e\beta L}{d} \left(\frac{d}{\hat{a}\beta} + |\theta^*|^2 \right) \right) = O\left(d \log \left(C^* \left(1 + \frac{\beta}{d} \right) \right) \right). \tag{4.42}$$

with $C^* > 0$ independent of β, d, n. Finally, one obtains

$$\mathbb{E}[\hat{U}(\hat{\theta})] - \inf_{\theta \in \mathbb{R}^d} \hat{U}(\theta) \leq \hat{C}_5 e^{-\hat{C}_4 \gamma n} + \hat{C}_6 \sqrt{\gamma} + \hat{C}_7 / \beta.$$

4.5 Application

4.5.1 Quantile estimation with $\mathscr{L}^2$ regularization

We consider the quantile estimation for AR(1) process with $\mathscr{L}^2$ regularization, which has been discussed in [8], [28], [49]. The data $Z_t \in \mathbb{R}$, $t \in \mathbb{Z}$, follows an AR(1) process given by

$$Z_{t+1} = \tilde{\alpha} Z_t + \bar{\xi}_{t+1},$$

where $\tilde{\alpha}$ is a constant with $|\tilde{\alpha}| < 1$ and $(\bar{\xi}_t)_{t \in \mathbb{Z}}$ are i.i.d. standard normal random variables. The above expression can be further rewritten as

$$Z_t = \sum_{j=0}^{\infty} \tilde{\alpha}^j \bar{\xi}_{t-j}.$$

One notes that Z_t has a stationary distribution π_Z which is normally distributed with mean 0 and variance $1/(1 - \tilde{\alpha}^2)$. Our task is to identify the q-th quantile of the stationary distribution π_Z using the SGLD algorithm (4.1), in other words, we aim to solve the following problem: for any $c > 0$,

$$\min_\theta \mathbb{E}\left[l_q(Z_\infty - \theta)\right] + c|\theta|^2,$$

where $Z_\infty \sim \pi_Z$ and

$$l_q(x) = \begin{cases} qx, & x \geq 0, \\ (q-1)x, & x < 0. \end{cases}$$

The stochastic gradient $H : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ of the SGLD algorithm (4.1) is given by

$$H(\theta, z) = -q + \mathbb{1}_{\{z < \theta\}} + 2c\theta, \tag{4.43}$$

where c is a positive constant. To check that **C-1** is satisfied, denote by $F(\theta, z) = -q + 2c\theta$, $G(\theta, z) = \mathbb{1}_{\{z < \theta\}}$. It can be easily seen that **C-1** holds with $\rho = 0$, $L_1 = 2c$, $L_2 = 0$ and $K_1(z) = 1$. Then, by Remark 4.3 and its proof in Appendix C.1, assumption **C-3** holds with $L = 2c + 1$. Moreover, **C-4** holds with $A(z) = c\mathbf{I}_d$ and $b(z) = q^2/(4c)$, which implies $a = c$ and

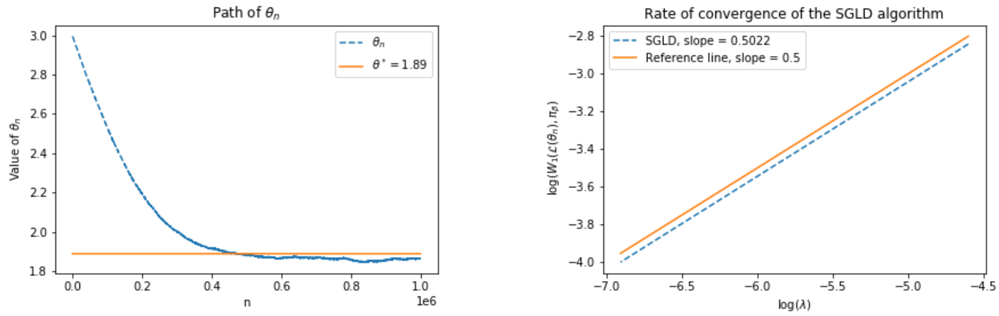


Figure 4.5.1: *[Left] Path of θ_n when $q = 0.95$. [Right] Rate of convergence of the SGLD algorithm.*

$b = q^2/(4c)$.

One notes that the value of the q-th quantile of π_Z is given by $\theta^* = N(q)/\sqrt{1-\tilde{\alpha}^2}$ where $N(\cdot)$ is the cumulative distribution function of the standard normal distribution. For the simulation, set $\tilde{\alpha} = 0.5$, $q = 0.95$, and thus, $\theta^* = 1.89$. Moreover, let $\theta_0 = 3$, $\beta = 10^8$, $c = 10^{-6}$. Note that we use the step restriction given in Remark 4.19 for all the examples in this section. In Figure 4.5.1, the left graph is obtained by using the SGLD algorithm (4.1) with $\gamma = 10^{-4}$ and the number of iterations $n = 10^6$. It shows the path of θ_n with the first 10000 iterations being discarded, and the path stabilises at around the true value $\theta^* = 1.89$. The right graph of Figure 4.5.1 illustrates the rate of convergence of the SGLD algorithm in Wasserstein-1 distance based on 5000 samples. The slope of the results in W_1 obtained using numerical experiments is 0.5022, which supports our theoretical finding in Theorem 4.6 with rate 1/2. One notes that the samples from π_β is generated by running the SGLD algorithm with $\gamma = 10^{-5}$ and $n = 10^7$.

4.5.2 Modified Kohonen algorithm

We consider the vector quantization using Kohonen algorithm, see e.g. [4], [18]. The aim is to obtain the optimal quantizer for a one-dimensional random variable Z. For $\theta := (\theta^1, \ldots, \theta^N)$ with $1 \le i \le N$, the Voronoi cells are defined as

$$\mathcal{V}^i(\theta) := \left\{ z \in \mathbb{R} : |z - \theta^i| = \min_{1 \le j \le N} |z - \theta^j| \right\}$$

The zero-neighbourhood Kohonen algorithm is proposed to minimize the following quantity in θ:

$$\sum_{i=1}^{N} \left(\mathbb{E}\left[|Z - \theta^i|^2 \mathbb{1}_{\mathcal{V}^i(\theta)}(Z) \right] + c|\theta^i|^2 \right),$$

and it can be written explicitly as, for any $n \in \mathbb{N}$ and for every $i = 1, \ldots, N$,

$$\theta^i_{n+1} = \theta^i_n + 2\gamma \left(\mathbb{1}_{\mathcal{V}^i(\theta_n)}(\bar{Z}_{n+1})(\bar{Z}_{n+1} - \theta^i_n) - c\theta^i_n \right) + \sqrt{2\beta^{-1}\gamma}\bar{\xi}_{n+1},$$

where $\bar{Z}$ is a process whose stationary distribution coincides with the law of Z, and $(\bar{\xi}_n)_{n\in\mathbb{N}}$ are i.i.d. standard Normal random variables in $\mathbb{R}$. However, one notices that assumption **C-1** does not hold, as in this case, $G(\theta^i, z) = 2\mathbb{1}_{\mathcal{V}^i(\theta)}(z)(z - \theta^i)$ which is unbounded in both variables. To tackle this problem, we consider instead the following:

$$\sum_{i=1}^{N} \left(\mathbb{E}\left[(|Z - \theta^i|^2 \wedge R) \mathbb{1}_{\mathcal{V}^i(\theta)}(Z) \right] + c|\theta^i|^2 \right),$$

where $R > 0$ is a sufficiently large value. Then, the stochastic gradient $\tilde{H} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ can be expressed as

$$\tilde{H}(\theta^i, z) = -2 \left((z - \theta^i) \mathbb{1}_{\left\{-\sqrt{R}+\theta^i \le z \le \sqrt{R}+\theta^i\right\}} \mathbb{1}_{\mathcal{V}^i(\theta)}(z) - c\theta^i \right).$$

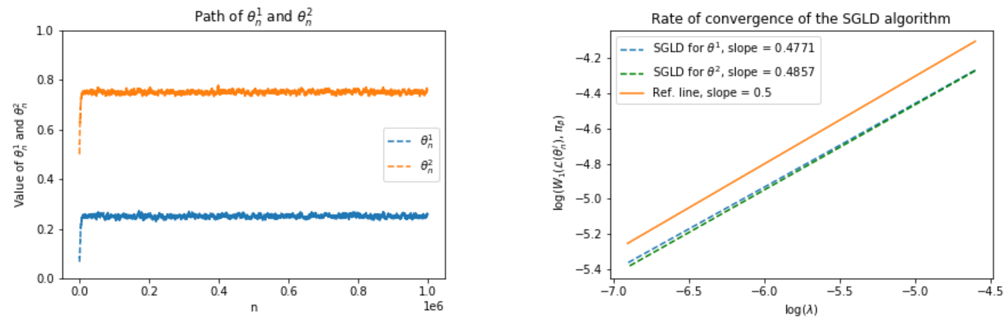


Figure 4.5.2: *[Left] Path of θ_n^1 and θ_n^2. [Right] Rate of convergence of the SGLD algorithm based on 10000 samples.*

Denote by $\tilde{F}(\theta^i, z) = 2c\theta^i$ and $\tilde{G}(\theta^i, z) = -2(z-\theta^i)\mathbb{1}_{\{-\sqrt{R}+\theta^i \le z \le \sqrt{R}+\theta^i\}}\mathbb{1}_{\mathcal{V}^i(\theta)}(z)$. Then, one can check that **C-1** holds with $\rho = 0$, $L_1 = 2c$, $L_2 = 0$ and $K_1(z) = 2\sqrt{R}$. As for **C-3**, one considers $\mathcal{V}^i(\theta)$ takes the form of $(-\infty, \bar{g}(\theta))$, where $\bar{g}$ is Lipschitz continuous with Lipschitz constant L_g and we assume $|z - \bar{\theta}^i| \le |z - \theta^i|$ without loss of generality. Then, one obtains

$$
\begin{aligned}
&\mathbb{E}\left[\left|\tilde{H}(\theta^i, Z) - \tilde{H}(\bar{\theta}^i, Z)\right|\right] \\
&\le 2c|\theta^i - \bar{\theta}^i| + 2\mathbb{E}\Big[\Big|(Z-\theta^i)\mathbb{1}_{\{-\sqrt{R}+\theta^i \le Z \le \sqrt{R}+\theta^i\}}\mathbb{1}_{\mathcal{V}^i(\theta)}(Z) \\
&\qquad\qquad\qquad\qquad -(Z-\bar{\theta}^i)\mathbb{1}_{\{-\sqrt{R}+\bar{\theta}^i \le Z \le \sqrt{R}+\bar{\theta}^i\}}\mathbb{1}_{\mathcal{V}^i(\bar{\theta})}(Z)\Big|\Big] \\
&\le 2c|\theta^i - \bar{\theta}^i| + 2\mathbb{E}\Big[\Big|(Z-\theta^i)\mathbb{1}_{\{-\sqrt{R}+\theta^i \le Z \le \sqrt{R}+\theta^i\}}\mathbb{1}_{\mathcal{V}^i(\theta)}(Z) \\
&\qquad\qquad\qquad\qquad -(Z-\theta^i)\mathbb{1}_{\{-\sqrt{R}+\theta^i \le Z \le \sqrt{R}+\theta^i\}}\mathbb{1}_{\mathcal{V}^i(\bar{\theta})}(Z)\Big|\Big] \\
&\quad + 2\mathbb{E}\Big[\Big|(Z-\theta^i)\mathbb{1}_{\{-\sqrt{R}+\theta^i \le Z \le \sqrt{R}+\theta^i\}}\mathbb{1}_{\mathcal{V}^i(\bar{\theta})}(Z) \\
&\qquad\qquad\qquad\qquad -(Z-\bar{\theta}^i)\mathbb{1}_{\{-\sqrt{R}+\theta^i \le Z \le \sqrt{R}+\theta^i\}}\mathbb{1}_{\mathcal{V}^i(\bar{\theta})}(Z)\Big|\Big] \\
&\quad + 2\mathbb{E}\Big[\Big|(Z-\bar{\theta}^i)\mathbb{1}_{\{-\sqrt{R}+\theta^i \le Z \le \sqrt{R}+\theta^i\}}\mathbb{1}_{\mathcal{V}^i(\bar{\theta})}(Z) \\
&\qquad\qquad\qquad\qquad -(Z-\bar{\theta}^i)\mathbb{1}_{\{-\sqrt{R}+\bar{\theta}^i \le Z \le \sqrt{R}+\bar{\theta}^i\}}\mathbb{1}_{\mathcal{V}^i(\bar{\theta})}(Z)\Big|\Big] \\
&\le 2c|\theta^i - \bar{\theta}^i| + 2\sqrt{R}c_d L_g|\theta^i - \bar{\theta}^i| + 2|\theta^i - \bar{\theta}^i| \\
&\quad + 2\mathbb{E}\left[\left|\mathbb{1}_{\{-\sqrt{R}+\theta^i \le Z \le \sqrt{R}+\theta^i\}} - \mathbb{1}_{\{-\sqrt{R}+\bar{\theta}^i \le Z \le \sqrt{R}+\bar{\theta}^i\}}\right|\right](\sqrt{R} + |\theta^i - \bar{\theta}^i|) \\
&\le 2c|\theta^i - \bar{\theta}^i| + 2\sqrt{R}c_d L_g|\theta^i - \bar{\theta}^i| + 2|\theta^i - \bar{\theta}^i| \\
&\quad + 2\mathbb{E}\left[\left|\mathbb{1}_{\{-\sqrt{R}+\theta^i \le Z \le -\sqrt{R}+\bar{\theta}^i\}} + \mathbb{1}_{\{\sqrt{R}+\theta^i \le Z \le \sqrt{R}+\bar{\theta}^i\}}\right|\right](\sqrt{R} + |\theta^i - \bar{\theta}^i|) \\
&\le 2c|\theta^i - \bar{\theta}^i| + 2\sqrt{R}c_d L_g|\theta^i - \bar{\theta}^i| + 2|\theta^i - \bar{\theta}^i| + 4(2\sqrt{R}c_d + 1)|\theta^i - \bar{\theta}^i| \\
&\le 2\left(c + \sqrt{R}c_d L_g + 1 + 2(2\sqrt{R}c_d + 1)\right)|\theta^i - \bar{\theta}^i|,
\end{aligned}
\tag{4.44}
$$

where c_d is the upper bound of the density of Z and the proof for the third inequality above can be found in Appendix C.3. Thus, assumption **C-3** holds with $2(c+\sqrt{R}c_d L_g+1+2(2\sqrt{R}c_d+1))$. One notices that similar arguments can be applied when $\mathcal{V}^i(\theta)$ takes the form of $(\bar{g}(\theta), \infty)$ or $(\bar{g}(\theta), \tilde{g}(\theta))$ with $\bar{g}$, $\tilde{g}$ being Lipschitz continuous. Moreover, assumption **C-4** holds with $A(z) = 2c\mathbf{I}_d$ and $b(z) = 0$, which implies $a = 2c$ and $b = 0$.

For the numerical experiments, consider $N = 2$, $\theta = (\theta^1, \theta^2)$, $\mathcal{V}^1(\theta) = (0, (\theta^1 + \theta^2)/2]$ and $\mathcal{V}^2(\theta) = [(\theta^1 + \theta^2)/2, 0)$. The data sequence $(\bar{Z}_n)_{n\in\mathbb{N}}$ are i.i.d. observations from the uniform distribution on $[0, 1]$, which is a classical case studied in literature, see [4] and references therein. Set $\theta_0^1 = 0.05$, $\theta_0^2 = 0.03$, $\beta = 10^8$, $c = 10^{-6}$ and the number of iterations $n = 10^6$. In this case, the optimal values for θ^1 and θ^2 are $1/4$ and $3/4$, respectively, which are supported by the

	$\bar{q} = 0.95$				$\bar{q} = 0.99$			
	VaR*	CVaR*	$\mathrm{VaR}_{\mathrm{SGLD}}$	$\mathrm{CVaR}_{\mathrm{SGLD}}$	VaR*	CVaR*	$\mathrm{VaR}_{\mathrm{SGLD}}$	$\mathrm{CVaR}_{\mathrm{SGLD}}$
$\mu = 0, \sigma = 1$	1.645	2.062	1.642 (0.02)	2.062 (0.0006)	2.326	2.677	2.329 (0.04)	2.662 (0.0038)
$\mu = 1, \sigma = 2$	4.290	5.124	4.294 (0.03)	5.126 (0.0006)	5.653	6.335	5.640 (0.06)	6.336 (0.0032)
$\mu = 3, \sigma = 5$	11.224	13.311	11.230 (0.05)	13.305 (0.0006)	14.632	16.337	14.643 (0.11)	16.313 (0.006)

Table 4.1: VaR and CVaR for normal distribution $N(\mu, \sigma)$.

numerical approximations in Figure 4.5.2. Moreover, as illustrated, the rates of convergence of the SGLD algorithm for θ^1 and θ^2 are consistent with the theoretical results in Theorem 4.6.

4.5.3 VaR-CVaR algorithm

In this section, we consider the problem of computing Value-at-Risk (VaR) and Conditional-Value-at-Risk (CVaR), which are commonly used risk measures in financial risk management. In order to obtain the two quantities, one considers the following optimization problem:

$$\min_{\theta} V(\theta) = \min_{\theta} \left(\mathbb{E}\left[\theta + \frac{1}{1-\bar{q}} (f(Z) - \theta)_+ \right] + c|\theta|^2 \right), \tag{4.45}$$

where $0 < \bar{q} < 1$, f is continuous and $f(Z)$ is integrable with respect to the probability measure. Then, by [2, Proposition 2.1], $\mathrm{VaR}_{\bar{q}}(f(Z)) = \operatorname{argmin} V(\theta)$ and $\mathrm{CVaR}_{\bar{q}}(f(Z)) = \min_\theta V(\theta)$.

Single asset: First, we consider minimizing CVaR for a single asset. Let $f(z) = z$. To compute VaR, the stochastic gradient $H : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ of the SGLD algorithm (4.1) is given by

$$H(\theta, z) = 1 - \frac{1}{1-\bar{q}} \mathbb{1}_{\{f(z) \geq \theta\}} + 2c\theta = -\frac{\bar{q}}{1-\bar{q}} + \frac{1}{1-\bar{q}} \mathbb{1}_{\{f(z) < \theta\}} + 2c\theta.$$

One notices that the above expression has a similar form as (4.43). Then, one can check that assumptions **C-1** - **C-4** are satisfied. More precisely, denote by $F(\theta, z) = -\bar{q}/(1-\bar{q}) + 2c\theta$, $G(\theta, z) = \mathbb{1}_{\{z<\theta\}}/(1-\bar{q})$, one observes that **C-1** holds with $\rho = 0$, $L_1 = 2c$, $L_2 = 0$ and $K_1(z) = 1/(1-\bar{q})$. Let Z be a one-dimensional random variable with fourth absolute moment, then assumption **C-2** is satisfied. Denote by $\bar{c}_d$ the upper bound of the density of Z, assumption **C-3** holds with $L = 2c + \bar{c}_d/(1-\bar{q})$. Furthermore, assumption **C-4** holds with $A(z) = c\mathbf{I}_d$ and $b(z) = \bar{q}^2/(4c(1-\bar{q})^2)$, which implies $a = c$ and $b = \bar{q}^2/(4c(1-\bar{q})^2)$.

For the numerical experiments, we set $\theta_0 = 0$, $\beta = 10^8$, $c = 10^{-8}$, $\gamma = 10^{-4}$ and the number of iterations $n = 10^6$. Table 4.1 and 4.2 present VaR and CVaR for the normal distribution and Student's t-distribution. VaR* and CVaR* in the tables denote the theoretical values, while $\mathrm{VaR}_{\mathrm{SGLD}}$ and $\mathrm{CVaR}_{\mathrm{SGLD}}$ denote the numerical approximations from the SGLD algorithm (4.1). Each approximation in the table is obtained based on 10000 samples, which is followed by its sample standard deviation shown in brackets. In addition, in Figure 4.5.3, the left graph illustrates the path of θ_n for the t-distribution, whereas the right graph shows that the rate of convergence of the SGLD algorithm (4.1) is 0.4811. One notes that the samples from π_β is generated by running the SGLD algorithm with $\gamma = 10^{-5}$, and $n = 10^7$. Furthermore, in the case that the data process $(Z_n)_{n\in\mathbb{N}}$ consists of i.i.d. observations from Student's t-distribution with d.f.= 10, Figure 4.5.4 shows the path of the expected excess risk, i.e.

$$\mathbb{E}[V(\theta_n^\gamma)] - V(\theta^*),$$

where $V(\theta^*) = \inf_{\theta\in\mathbb{R}} V(\theta)$ and $\theta^* = 1.8125$.

Portfolios: To minimize CVaR for a given portfolio, we consider the following optimization problem:

$$\min_{\hat{\theta}} V(\hat{\theta}) = \min_{\hat{\theta}} \left(\mathbb{E}\left[\frac{1}{1-\bar{q}} \left(\sum_{i=1}^n g_i(w) Z_i - \theta \right)_+ + \theta \right] + c|\hat{\theta}|^2 \right), \tag{4.46}$$

where the parameter $\hat{\theta} := (\theta, w)^\intercal = (\theta, w_1, \ldots, w_n)^\intercal$ and $g_i(w) := \frac{e^{w_i}}{\sum_{j=1}^n e^{w_j}}$ for $i = 1, \ldots, n$.

	$\bar{q} = 0.95$				$\bar{q} = 0.99$			
	VaR*	CVaR*	$\mathrm{VaR}_{\mathrm{SGLD}}$	$\mathrm{CVaR}_{\mathrm{SGLD}}$	VaR*	CVaR*	$\mathrm{VaR}_{\mathrm{SGLD}}$	$\mathrm{CVaR}_{\mathrm{SGLD}}$
d.f. = 10	1.812	2.416	1.808 (0.02)	2.407 (0.0005)	2.764	3.357	2.767 (0.05)	3.350 (0.003)
d.f. = 7	1.895	2.595	1.895 (0.03)	2.594 (0.0008)	2.998	3.757	3.001 (0.05)	3.782 (0.0024)
d.f. = 3	2.353	3.876	2.358 (0.03)	3.873 (0.0008)	4.541	6.968	4.542 (0.08)	6.967 (0.0028)

Table 4.2: VaR and CVaR for Student's t distribution.

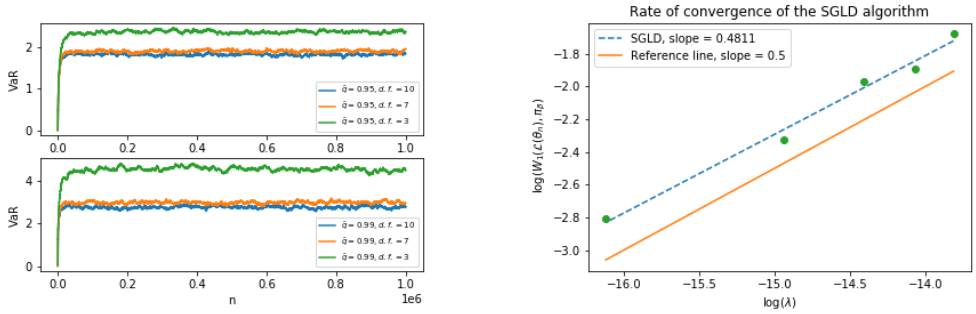


Figure 4.5.3: *[Left] Path of θ_n (VaR) for Student's t-distribution. [Right] Rate of convergence of the SGLD algorithm based on 5000 samples.*

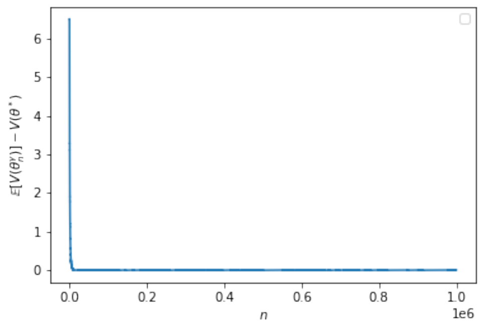


Figure 4.5.4: *Path of the expected excess risk of V w.r.t. n.*

By solving (4.46), we obtain not only VaR for a given portfolio, but also the optimal weight for each asset in the portfolio such that CVaR is minimized. For $i = 1, \ldots, n$, let Z_i's be i.i.d. one-dimensional random variables with fourth moment, and denote by c_Z, $c_{\bar{Z}}$ their first and second absolute moment respectively. Moreover, denote by f_{Z_i} the density function of Z_i for any $i = 1, \ldots, n$, and we further assume $|z| f_{Z_i}(z)$ is bounded for any i and $z \in \mathbb{R}$. Note that the assumption is satisfied for a wide range of distributions, for example, the distributions shown in Table 4.3.

The stochastic gradient $H_{\hat{\theta}}(\hat{\theta}, z) : \mathbb{R}^{n+1} \times \mathbb{R}^n \to \mathbb{R}^{n+1}$ is defined as

$$H_{\hat{\theta}}(\hat{\theta}, z) := (H_\theta(\hat{\theta}, z), H_{w_1}(\hat{\theta}, z), \ldots, H_{w_n}(\hat{\theta}, z))^\intercal,$$

where $H_\theta(\hat{\theta}, z) : \mathbb{R}^{n+1} \times \mathbb{R}^n \to \mathbb{R}$ and $H_{w_j}(\hat{\theta}, z) : \mathbb{R}^{n+1} \times \mathbb{R}^n \to \mathbb{R}$ for all j are given by

$$H_\theta(\hat{\theta}, z) = 1 - \frac{1}{1-\bar{q}} \mathbb{1}_{\{\sum_{i=1}^n g_i(w) z_i \geq \theta\}} + 2c\theta,$$

and

$$H_{w_j}(\hat{\theta}, z) = \frac{1}{1-\bar{q}} \hat{g}_{w_j}(w, z) \mathbb{1}_{\{\sum_{i=1}^n g_i(w) z_i \geq \theta\}} + 2cw_j,$$

where

$$\hat{g}_{w_j}(w,z) = \sum_{i=1}^{n} \frac{\partial g_i(w)}{\partial w_j} z_i$$

for any $j = 1, \ldots, n$ with $\frac{\partial g_j(w)}{\partial w_j} = \frac{e^{w_j}(\sum_{l\neq j} e^{w_l})}{(\sum_{l=1}^n e^{w_l})^2}$, and $\frac{\partial g_i(w)}{\partial w_j} = -\frac{e^{w_i}e^{w_j}}{(\sum_{l=1}^n e^{w_l})^2}$ for $i \neq j$. One notes that $|\hat{g}_{w_j}(w,z)| \leq \sum_{i=1}^n |z_i|$ for any j.

To see assumptions **C-1** - **C-4** hold for $H_{\hat{\theta}}(\hat{\theta}, z)$, we first show that the assumptions hold for H_θ. Denote by

$$F_\theta(\hat{\theta}, z) = 2c\theta, \quad G_\theta(\hat{\theta}, z) = 1 - \mathbb{1}_{\{\sum_{i=1}^n g_i(w)z_i \geq \theta\}}/(1-\bar{q}),$$

then $H_\theta = F_\theta + G_\theta$. Assumption **C-1** holds with $\rho = 0$, $L_1 = 2c$, $L_2 = 0$ and $K_1(z) = (2-\bar{q})/(1-\bar{q})$. By taking into consideration the expression of $K_1(z)$ and the construction of the problem, **C-2** is satisfied. Assumption **C-4** holds with $A(z) = 2c\mathbf{I}_d$ and $b(z) = 0$, which implies $a = 2c$ and $b = 0$. To check assumption **C-3**, one considers $\hat{\theta}' := (\bar{\theta}, w)^\intercal$, and calculates by assuming without loss of generality $g_n(w) = \max\{g_1(w), \ldots, g_n(w)\}$

$$\begin{aligned}
&\mathbb{E}\left[\left|H_\theta(\hat{\theta}, Z) - H_\theta(\hat{\theta}', Z)\right|\right] \\
&\leq 2c\left|\theta - \bar{\theta}\right| + \frac{1}{1-\bar{q}}\mathbb{E}\left[\left|\mathbb{1}_{\{\sum_{i=1}^n g_i(w)Z_i \geq \theta\}} - \mathbb{1}_{\{\sum_{i=1}^n g_i(w)Z_i \geq \bar{\theta}\}}\right|\right] \\
&\leq 2c\left|\hat{\theta} - \hat{\theta}'\right| + \frac{1}{1-\bar{q}}(E_1 + E_2),
\end{aligned}$$

where

$$E_1 = \mathbb{E}\left[\mathbb{1}_{\{\theta \leq \sum_{i=1}^n g_i(w)Z_i \leq \bar{\theta}\}}\right], \quad E_2 = \mathbb{E}\left[\mathbb{1}_{\{\bar{\theta} \leq \sum_{i=1}^n g_i(w)Z_i \leq \theta\}}\right].$$

To estimate E_1, one writes

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{\{\theta \leq \sum_{i=1}^n g_i(w)Z_i \leq \bar{\theta}\}}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\{(\theta - \sum_{i\neq n} g_i(w)Z_i)/g_n(w)) \leq Z_n \leq (\bar{\theta} - \sum_{i\neq n} g_i(w)Z_i)/g_n(w))\}}\middle| Z_1, \ldots, Z_{n-1}\right]\right] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{(\theta - \sum_{i\neq n} g_i(w)z_i)/g_n(w)}^{(\bar{\theta} - \sum_{i\neq n} g_i(w)z_i)/g_n(w)} f_{Z_n}(y)\, dy f_{Z_{n-1}}(z_{n-1})\, dz_{n-1} \cdots f_{Z_1}(z_1)\, dz_1 \\
&\leq nc_{Z_n}\left|\hat{\theta} - \hat{\theta}'\right|,
\end{aligned}$$

where we use the fact $g_n(w) \geq 1/n$ in the last inequality and c_{Z_n} denotes the upper bound of the density of Z_n. E_2 can be estimated by using similar arguments. Then, one obtains

$$\mathbb{E}\left[\left|H_\theta(\hat{\theta}, Z) - H_\theta(\hat{\theta}', Z)\right|\right] \leq (2c + 2nc_{Z_n}/(1-\bar{q}))\left|\hat{\theta} - \hat{\theta}'\right|,$$

which implies **C-3** holds with $L = 2c + 2nc_{Z_n}/(1-\bar{q})$.

Next, we check assumptions for H_{w_j}. Denote by

$$F_{w_j}(\hat{\theta}, z) = 2cw_j, \quad G_{w_j}(\hat{\theta}, z) = \hat{g}_{w_j}(w,z)\mathbb{1}_{\{\sum_{i=1}^n g_i(w)z_i \geq \theta\}}/(1-\bar{q}),$$

then $H_{w_j} = F_{w_j} + G_{w_j}$. Assumption **C-1** holds with $\rho = 0$, $L_1 = 2c$, $L_2 = 0$ and $K_1(z) = \sum_i |z_i|/(1-\bar{q})$. By taking into consideration the expression of $K_1(z)$ and the construction of the problem, **C-2** is satisfied. Assumption **C-4** holds with $A(z) = 2c\mathbf{I}_d$ and $b(z) = 0$, which implies $a = 2c$ and $b = 0$. Then, we check **C-3** for H_{w_1}, and the arguments stay the same lines for any other H_{w_j}, $j = 2, \ldots, n$. Consider $\hat{\theta}^\sharp := (\theta, \bar{w})^\intercal = (\theta, \bar{w}_1, w_2, \ldots, w_n)^\intercal$. One calculates

$$\begin{aligned}
&\mathbb{E}\left[\left|H_{w_1}(\hat{\theta}, Z) - H_{w_1}(\hat{\theta}^\sharp, Z)\right|\right] \\
&\leq 2c\left|w_1 - \bar{w}_1\right| + \frac{1}{1-\bar{q}}\mathbb{E}\left[\left|\hat{g}_{w_1}(w,Z)\mathbb{1}_{\{\sum_{i=1}^n g_i(w)Z_i \geq \theta\}} - \hat{g}_{w_1}(\bar{w},Z)\mathbb{1}_{\{\sum_{i=1} g_i(\bar{w})Z_i \geq \theta\}}\right|\right]
\end{aligned}$$

$$
\begin{aligned}
&\leq 2c\left|\hat{\theta}-\hat{\theta}^{\sharp}\right|+\frac{1}{1-\bar{q}}\mathbb{E}\left[\left|\hat{g}_{w_1}(w,Z)\mathbb{1}_{\{\sum_{i=1}^n g_i(w)Z_i\geq\theta\}}-\hat{g}_{w_1}(\bar{w},Z)\mathbb{1}_{\{\sum_{i=1}^n g_i(w)Z_i\geq\theta\}}\right|\right] \\
&\quad+\frac{1}{1-\bar{q}}\mathbb{E}\left[\left|\hat{g}_{w_1}(\bar{w},Z)\mathbb{1}_{\{\sum_{i=1}^n g_i(w)Z_i\geq\theta\}}-\hat{g}_{w_1}(\bar{w},Z)\mathbb{1}_{\{\sum_{i=1} g_i(\bar{w})Z_i\geq\theta\}}\right|\right] \\
&\leq 2c\left|\hat{\theta}-\hat{\theta}^{\sharp}\right|+\frac{2nc_Z}{1-\bar{q}}\left|\hat{\theta}-\hat{\theta}^{\sharp}\right| \\
&\quad+\frac{1}{1-\bar{q}}\mathbb{E}\left[\left|\hat{g}_{w_1}(\bar{w},Z)\mathbb{1}_{\{\sum_{i=1}^n g_i(w)Z_i\geq\theta\}}-\hat{g}_{w_1}(\bar{w},Z)\mathbb{1}_{\{\sum_{i=1} g_i(\bar{w})Z_i\geq\theta\}}\right|\right],
\end{aligned}
$$

where the third inequality holds due to the fact that $|\hat{g}_{w_1}(w,Z)-\hat{g}_{w_1}(\bar{w},Z)|\leq 2|w_1-\bar{w}_1|\sum_i|Z_i|$, see Appendix C.5 for the proof. Then, by using $|\hat{g}_{w_1}(\bar{w},z)|\leq\sum_i|z_i|$,

$$
\begin{aligned}
&\mathbb{E}\left[\left|H_{w_1}(\hat{\theta},Z)-H_{w_1}(\hat{\theta}^{\sharp},Z)\right|\right] \\
&\leq 2c\left|\hat{\theta}-\hat{\theta}^{\sharp}\right|+\frac{2nc_Z}{1-\bar{q}}\left|\hat{\theta}-\hat{\theta}^{\sharp}\right| \\
&\quad+\frac{1}{1-\bar{q}}\mathbb{E}\left[\sum_i|Z_i|\left|\mathbb{1}_{\{\sum_{i=1}^n g_i(w)Z_i\geq\theta\}}-\mathbb{1}_{\{\sum_{i=1} g_i(\bar{w})Z_i\geq\theta\}}\right|\right] \\
&\leq(2c+2nc_Z/(1-\bar{q}))\left|\hat{\theta}-\hat{\theta}^{\sharp}\right| \\
&\quad+2(n-1)(c_Z(\bar{c}_{Z_n}+\bar{c}_{Z_1})+(c_{\bar{Z}}+(n-2)c_{\bar{Z}}^2)(c_{Z_n}+c_{Z_1}))/(1-\bar{q})\left|\hat{\theta}-\hat{\theta}^{\sharp}\right|,
\end{aligned}
\tag{4.47}
$$

where c_Z, $c_{\bar{Z}}$ denote the first and the second absolute moment of Z_i's respectively, $\bar{c}_{Z_i}$ is the upper bound of the function $|z|f_{Z_i}(z)$, and c_{Z_i} is the upper bound of the density of Z_i. Detailed calculations to obtain the last inequality in (4.47) is given in Appendix C.5. Thus assumption **C-3** holds with $L=2c+6nc_Z/(1-\bar{q})+2(n-1)(c_Z(\bar{c}_{Z_n}+\bar{c}_{Z_1})+(c_{\bar{Z}}+(n-2)c_{\bar{Z}}^2)(c_{Z_n}+c_{Z_1}))/(1-\bar{q})$.

For the numerical experiments, we set $\theta_0=0$, $\beta=10^8$, $c=10^{-8}$, $\gamma=10^{-4}$ and the number of iterations $n=10^6$. Table 4.3 illustrates 95% VaR and CVaR obtained using the SGLD algorithm for a portfolio of two assets Z_1 and Z_2 with weights $g_1(w)$ and $g_2(w)$ respectively. The reference values $g_1(w^*)$, $g_2(w^*)$, VaR* and CVaR* are obtained numerically in the following way, which is computationally expensive.

(i) First, we create 100 evenly spaced numbers over the interval $[0,1]$.

(ii) Then, for any given distributions of Z_1 and Z_2, assign each of the 100 numbers to $g_1(w)$, which is the weight of Z_1, and calculate the 95% CVaR for the combination $g_1(w)Z_1+g_2(w)Z_2$.

(iii) Finally, we obtain the minimum CVaR and the corresponding $g_1(w)$ among the 100 values and denote them as CVaR* and $g_1(w^*)$, here, one notes that the corresponding VaR* can be calculated using the optimal weights $g_1(w^*)$ and $g_2(w^*)$.

Moreover, Figure 4.5.5 shows that the rate of convergence of the SGLD algorithm (4.1) for the parameter w_1 is 0.5319, which supports the theoretical finding in Theorem 4.6. One notes that the samples from π_β is generated by running the SGLD algorithm with $\gamma=10^{-5}$, and $n=10^7$.

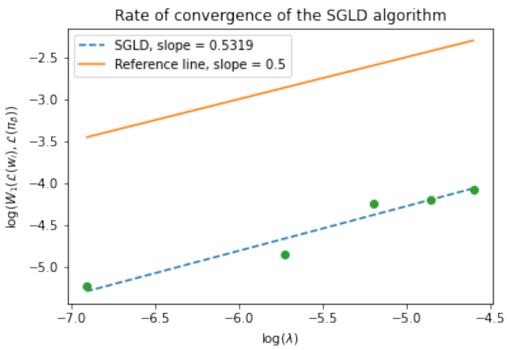


Figure 4.5.5: *Rate of convergence of the SGLD algorithm for w_1 based on 5000 samples.*

		SGLD algorithm				Reference			
Z_1	Z_2	$g_1(w)$	$g_2(w)$	$g_1(w)Z_1 + g_2(w)Z_2$		$g_1(w^*)$	$g_2(w^*)$	$g_1(w^*)Z_1 + g_2(w^*)Z_2$	
				$\mathrm{VaR}_{\mathrm{SGLD}}$	$\mathrm{CVaR}_{\mathrm{SGLD}}$			VaR*	CVaR*
$N(500, 1)$	$N(0, 10^{-4})$	0.00002	0.99998	0.025	0.03	0	1	0.016	0.021
$N(0, 10^6)$	$N(0, 10^{-4})$	0.000006	0.999994	0.016	0.25	0	1	0.016	0.021
$N(1, 4)$	$N(0, 1)$	0.111	0.889	1.615	2.004	0.11	0.89	1.617	1.999
$N(0, 1)$	t with d.f. $= 2.01$	0.917	0.083	1.567	1.975	0.9	0.1	1.531	1.971
$N(0, 1)$	t with d.f. $= 10$	0.577	0.423	1.236	1.554	0.58	0.42	1.224	1.553
$N(0, 1)$	t with d.f. $= 1000$	0.503	0.497	1.15	1.46	0.5	0.5	1.165	1.461
$N(1, 4)$	t with d.f. $= 2.01$	0.596	0.404	2.941	4.130	0.61	0.39	2.985	4.115
$N(1, 4)$	t with d.f. $= 10$	0.172	0.828	1.743	2.290	0.17	0.83	1.779	2.286
$N(1, 4)$	t with d.f. $= 1000$	0.113	0.887	1.594	2.008	0.11	0.89	1.619	2.002
$N(0, 1)$	Logistic(0,1)	0.775	0.225	1.422	1.816	0.78	0.22	1.442	1.813
$N(0, 1)$	Logistic(0,29)	0.999	0.001	1.633	2.110	1	0	1.645	2.063
$N(0, 1)$	Logistic(2,10)	0.997	0.003	1.650	2.101	1	0	1.648	2.065
$N(1, 4)$	Logistic(0,1)	0.402	0.598	2.635	3.262	0.4	0.6	2.607	3.261
$N(1, 4)$	Logistic(0,29)	0.998	0.002	4.284	5.145	1	0	4.284	5.116
$N(1, 4)$	Logistic(2,10)	0.991	0.009	4.255	5.132	0.99	0.01	4.283	5.114
$N(0, 1)$	Lognormal(0,1)	0.966	0.034	1.662	2.068	0.97	0.03	1.647	2.054
$N(0, 1)$	Lognormal(0,0.01)	0.074	0.926	1.145	1.205	0.07	0.93	1.132	1.186
$N(0, 1)$	Lognormal(1,4)	0.9997	0.0003	1.674	2.136	1	0	1.645	2.062
$N(1, 4)$	Lognormal(0,1)	0.732	0.268	3.750	4.605	0.74	0.26	3.771	4.599
$N(1, 4)$	Lognormal(0,0.01)	0.010	0.0.989	1.173	1.301	0	1	1.179	1.230
$N(1, 4)$	Lognormal(1,4)	0.997	0.003	4.266	5.194	1	0	4.292	5.129
Logistic(0,1)	Lognormal(0,1)	0.817	0.183	2.797	3.727	0.81	0.19	2.814	3.724
Logistic(0,1)	Lognormal(0,0.01)	0.022	0.978	1.169	1.256	0.02	0.98	1.164	1.217
Logistic(0,1)	Lognormal(1,4)	0.997	0.003	2.961	4.030	1	0	2.947	3.971
Logistic(2,10)	Lognormal(0,1)	0.043	0.956	5.245	8.412	0.04	0.96	5.198	8.400
Logistic(2,10)	Lognormal(0,0.01)	0.009	0.991	1.184	1.315	0	1	1.179	1.229
Logistic(2,10)	Lognormal(1,4)	0.996	0.004	31.651	41.748	0.99	0.01	31.420	41.738

Table 4.3: 95% VaR and CVaR for portfolios of two assets Z_1, Z_2 with the form $g_1(w)Z_1 + g_2(w)Z_2$.

Chapter 5

Future work

In this thesis, a tamed explicit order 1.5 scheme is proposed to approximate an SDE with super-linear coefficients. Then, it is applied to a Langevin SDE with super-linear drift coefficient. A higher order LMC algorithm is obtained which can be used in the sampling problem in statistical machine learning. It is thus a natural extension to consider the application of the order 1.5 scheme to the SGLD algorithm. One may first consider the case where the drift coefficient of the SDE is Lipschitz continuous. The convergence results in Wasserstein distances of such an algorithm can be established in both convex and non-convex setting. Then, one can extend the result to the case where the coefficient satisfies a local Lipschitz condition, and the taming technique can be applied. Moreover, in view of the Kohonen example in section 4.5, one may notice that it is of great importance to consider the case where the coefficient is locally Lipschitz and discontinuous.

Appendix A

Auxiliary results to Chapter 2

A.1 Proof of validity of the examples in Section 2.5

1. Consider the one-dimensional SDE

$$dx_t = x_t(1-x_t^2)dt + c(1-x_t^2)dw_t, \quad \forall t \in [0,T].$$

   (a) **A-1** is satisfied as x_0 is taken to be a constant (i.e. $x_0 = 3$).

   (b) To verify **A-2**, one calculates

$$\begin{aligned} 2xb(x) + (p_0-1)|\sigma(x)|^2 &= 2x^2 - 2x^4 + (p_0-1)c^2(1-x^2)^2 \\ &= (p_0-1)c^2 + 2(1-c^2(p_0-1))x^2 + (c^2(p_0-1)-2)x^4. \end{aligned}$$

   We require $c^2(p_0-1) - 2 \le 0$, which implies $p_0 \le \frac{2}{c^2} + 1$.

   (c) As for **A-3**, one writes

$$\begin{aligned} &2(x-\bar{x})(b(x)-b(\bar{x})) + (p_1-1)|\sigma(x)-\sigma(\bar{x})|^2 \\ &= 2(x-\bar{x})((x-x^3)-(\bar{x}-\bar{x}^3)) + (p_1-1)c^2|(1-x^2)-(1-\bar{x}^2)|^2 \\ &= 2(x-\bar{x})^2 - 2(x-\bar{x})^2((x+\bar{x})^2 - x\bar{x}) + (p_1-1)c^2|x+\bar{x}|^2|x-\bar{x}|^2 \\ &\le 2(x-\bar{x})^2 + (x-\bar{x})^2\left((p_1-1)c^2|x+\bar{x}|^2 - (x+\bar{x})^2\right). \end{aligned}$$

   Then, in order to guarantee $2(x-\bar{x})(b(x)-b(\bar{x}))+(p_1-1)|\sigma(x)-\sigma(\bar{x})|^2 \le K|x-\bar{x}|^2$ is satisifed for some $K > 0$, we require $p_1 \in (2, \frac{1}{c^2}+1]$.

   (d) The second derivative of $b(x) = x(1-x^2)$ is $-6x$, then **A-4** is satisfied with $\rho \ge 2$ since

$$\left|\frac{\partial^2 b(x)}{\partial x^2} - \frac{\partial^2 b(\bar{x})}{\partial \bar{x}^2}\right| \le 6|x-\bar{x}|$$

   (e) Similary, one can calculate the second derivative of $\sigma(x) = c(1-x^2)$, which is $-2c$. The assumption **A-5** is satisfied with $\rho \ge 2$.

   We choose ρ to be 2, then, since it is assumed in Theorem 1 that $p_0 \ge 2(5\rho+1) = 22$, one obtains $c \in [-0.3086, 0.3086]$ by using $p_0 \in [22, \frac{2}{c^2}+1]$ and $p_1 \in (2, \frac{1}{c^2}+1]$.

2. As for the second example, consider the one-dimensional SDE

$$dx_t = x_t(1-|x_t|^3)dt + c|x_t|^{\frac{5}{2}}dw_t, \quad \forall t \in [0,T].$$

   (a) We take $x_0 = 3$, therefore **A-1** is satisfied.

   (b) As for **A-2**, one calculates

$$2xb(x) + (p_0-1)|\sigma(x)|^2 = 2x^2 - 2|x|^5 + (p_0-1)c^2|x|^5$$

$$= 2x^2 + ((p_0 - 1)c^2 - 2)|x|^5.$$

To guarantee **A-2** is satisfied, we require $p_0 \leq \frac{2}{c^2} + 1$.

(c) To verify **A-3**, one calculates the following

$$\begin{aligned}
&2(x-\bar{x})(b(x)-b(\bar{x})) + (p_1-1)|\sigma(x)-\sigma(\bar{x})|^2 \\
&= 2(x-\bar{x})((x - x|x|^3) - (\bar{x} - \bar{x}|\bar{x}|^3)) + (p_1-1)c^2 \left| |x|^{\frac{5}{2}} - |\bar{x}|^{\frac{5}{2}} \right|^2 \\
&= 2(x-\bar{x})^2 - 2(|x|^5 - x\bar{x}|x|^3 - x\bar{x}|\bar{x}|^3 + |\bar{x}|^5) + (p_1-1)c^2 \left| |x|^{\frac{5}{2}} - |\bar{x}|^{\frac{5}{2}} \right|^2 \\
&\leq 2(x-\bar{x})^2 + \left(-2|x|^5 - 2|\bar{x}|^5 + \frac{6}{5}|x|^5 + \frac{6}{5}|\bar{x}|^5 + \frac{8}{5}|x|^{\frac{5}{2}}|\bar{x}|^{\frac{5}{2}} \right) \\
&\quad + (p_1-1)c^2 \left| |x|^{\frac{5}{2}} - |\bar{x}|^{\frac{5}{2}} \right|^2 \\
&= 2(x-\bar{x})^2 + \left((p_1-1)c^2 - \frac{4}{5} \right) \left| |x|^{\frac{5}{2}} - |\bar{x}|^{\frac{5}{2}} \right|^2.
\end{aligned}$$

Therefore, we require $p_1 \in (2, \frac{4}{5c^2} + 1]$ for **A-3** to be satisfied.

(d) The second derivative of $b(x) = x(1 - |x|^3)$ is $-12x|x|$, then **A-4** is satisfied with $\rho \geq 3$ since

$$\begin{aligned}
\left| \frac{\partial^2 b(x)}{\partial x^2} - \frac{\partial^2 b(\bar{x})}{\partial \bar{x}^2} \right| &\leq 12|\bar{x}|\bar{x}| - x|x|| \\
&= 12|\bar{x}|\bar{x}| - x|\bar{x}| + x|\bar{x}| - x|x|| \\
&\leq 12|\bar{x}||\bar{x} - x| + |x||\bar{x} - x| \\
&\leq 12(|x| + |\bar{x}|)|\bar{x} - x| \\
&\leq 12(1 + |x| + |\bar{x}|)|\bar{x} - x|.
\end{aligned}$$

(e) The second derivative of $\sigma(x) = c|x|^{\frac{5}{2}}$ is $\frac{15}{4}c|x|^{\frac{1}{2}}$, then one obtains

$$\left| \frac{\partial^2 \sigma(x)}{\partial x^2} - \frac{\partial^2 \sigma(\bar{x})}{\partial \bar{x}^2} \right| \leq \frac{15}{4}|c| \left| |x|^{\frac{1}{2}} - |\bar{x}|^{\frac{1}{2}} \right| \leq \frac{15}{4}|c||x - \bar{x}|^{\frac{1}{2}},$$

which implies that **A-5** is satisfied with $\rho \geq 4$, and the last inequality holds since

$$\left| |x|^{\frac{1}{2}} - |\bar{x}|^{\frac{1}{2}} \right|^2 \leq \left| |x|^{\frac{1}{2}} - |\bar{x}|^{\frac{1}{2}} \right| \left| |x|^{\frac{1}{2}} + |\bar{x}|^{\frac{1}{2}} \right| \leq ||x| - |\bar{x}|| \leq |x - \bar{x}|.$$

We choose $\rho = 4$, then, as it is assumed in Theorem 1 that $p_0 \geq 2(5\rho + 1) = 42$, one obtains $c \in [-0.2209, 0.2209]$ by using $p_0 \in [42, \frac{2}{c^2} + 1]$ and $p_1 \in (2, \frac{4}{5c^2} + 1]$.

Appendix B

Auxiliary results to Chapter 3

B.1 Proof of Remark 3.2

B-2 states there exists $L > 0$, $\rho \geq 2$, and $\alpha \in (0,1]$, such that for any $i = 1, \ldots, d$ and for all $x, y \in \mathbb{R}^d$,

$$|\nabla^2(\nabla U)^{(i)}(x) - \nabla^2(\nabla U)^{(i)}(y)| \leq L(1 + |x| + |y|)^{\rho-2}|x-y|^\alpha.$$

By **H2**, one obtains

$$|\nabla^2(\nabla U)^{(i)}(x)| \leq L(1+|x|)^{\rho-2}|x|^\alpha + |\nabla^2(\nabla U)^{(i)}(0)| \leq K(1+|x|)^{\rho-2+\alpha},$$

where $K = \max\{L, |\nabla^2(\nabla U)^{(i)}(0)|\}$. Then by fundamental theorem of calculus,

$$\begin{aligned}
|\nabla(\nabla U)^{(i)}(x) - \nabla(\nabla U)^{(i)}(y)| &= \left|\int_0^1 \nabla^2(\nabla U)^{(i)}(tx + (1-t)y)\, dt(x-y)\right| \\
&\leq \int_0^1 |\nabla^2(\nabla U)^{(i)}(tx + (1-t)y)|\, dt|x-y| \\
&\leq \int_0^1 K(1+|x|+|y|)^{\rho-2+\alpha}\, dt|x-y| \\
&\leq K(1+|x|+|y|)^{\rho-2+\alpha}|x-y|.
\end{aligned}$$

Moreover, notice that

$$\begin{aligned}
|\nabla^2 U(x) - \nabla^2 U(y)| &\leq |\nabla^2 U(x) - \nabla^2 U(y)|_{\mathsf{F}} \\
&= \left(\sum_{i=1}^d \sum_{j=1}^d \left|\frac{\partial^2 U(x)}{\partial x^{(i)}\partial x^{(j)}} - \frac{\partial^2 U(y)}{\partial x^{(i)}\partial x^{(j)}}\right|^2\right)^{1/2} \\
&= \left(\sum_{i=1}^d |\nabla(\nabla U)^{(i)}(x) - \nabla(\nabla U)^{(i)}(y)|^2\right)^{1/2} \\
&\leq \sqrt{d}K(1+|x|+|y|)^{\rho-2+\alpha}|x-y|.
\end{aligned}$$

Furthermore, one obtains

$$\begin{aligned}
&|\vec{\Delta}(\nabla U)(x) - \vec{\Delta}(\nabla U)(y)| \\
&= \left(\sum_{i=1}^d \left|\sum_{j=1}^d \frac{\partial^3 U(x)}{\partial x^{(i)}\partial x^{(j)}\partial x^{(j)}} - \frac{\partial^3 U(y)}{\partial x^{(i)}\partial x^{(j)}\partial x^{(j)}}\right|^2\right)^{1/2}
\end{aligned}$$

$$\leq \left(d \sum_{i=1}^{d} \sum_{j=1}^{d} \left| \frac{\partial^3 U(x)}{\partial x^{(i)} \partial x^{(j)} \partial x^{(j)}} - \frac{\partial^3 U(y)}{\partial x^{(i)} \partial x^{(j)} \partial x^{(j)}} \right|^2 \right)^{1/2}$$

$$\leq \left(d \sum_{i=1}^{d} |\nabla^2 (\nabla U)^{(i)}(x) - \nabla^2 (\nabla U)^{(i)}(y)|_{\mathsf{F}}^2 \right)^{1/2}$$

$$\leq \left(d^2 \sum_{i=1}^{d} L^2 (1 + |x| + |y|)^{2\rho - 4} |x - y|^{2\alpha} \right)^{1/2}$$

$$= d^{3/2} L (1 + |x| + |y|)^{\rho - 2} |x - y|^{\alpha}.$$

Notice that the last inequality in Remark 3.2 is not obtained directly by using the above result, but it is obtained by using the arguments in page 24 of [12]. However, the rest of the inequalities in Remark 3.2 can be obtained by using similar arguments as above.

B.2 Proof of inequality (3.10) in Proposition 3.10

In order to prove (3.10), one needs the following definition and the propositions.

Definition B.1. Consider a probability measure space $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d), \nu)$. Let $\mathcal{C}_L$ be the set of continuously differentiable, Lipschitz functions on $\mathbb{R}^d$. We say that ν satisfies a Log-Sobolev inequality if there exists $C > 0$ such that

$$\mathrm{Ent}_\nu(f^2) \leq 2C \int_{\mathbb{R}^d} |\nabla f|^2 \, d\nu,$$

for every function $f \in \mathcal{C}_L$ with $\mathrm{Ent}_\nu(f^2 log^+ f^2) < \infty$, where

$$\mathrm{Ent}_\nu(f) = \mathbb{E}_\nu(f \log f) - \mathbb{E}_\nu(f) \log \mathbb{E}_\nu(f).$$

For more details about the definition of the Log-Sobolev inequality, please refer to Chapter 2 in [31].

Proposition B.2 (Proposition 5.4.1 in [1])**.** *If ν satisfies a logarithmic Sobolev inequality with constant $C > 0$, then for every 1-Lipschitz function f and every $\tilde{\alpha}^2 < 1/C$,*

$$\int_{\mathbb{R}^d} e^{\tilde{\alpha}^2 f^2/2} \, d\nu < \infty.$$

More precisely, any 1-Lipschitz function f is integrable and for every $s \in \mathbb{R}$,

$$\int_{\mathbb{R}^d} e^{sf} \, d\nu < e^{s \int_{\mathbb{R}^d} f \, d\nu + Cs^2/2}.$$

Proposition B.3 (Proposition 5.5.1 in [1])**.** *The standard Gaussian measure ν on the Borel sets of $\mathbb{R}^d$ satisfies, for every $f \in \mathcal{C}_L$,*

$$\mathrm{Ent}_\nu(f^2) \leq 2 \int_{\mathbb{R}^d} |\nabla f|^2 \, d\nu.$$

Proposition B.3 implies that, for a Gaussian measure ν with mean μ and covariance matrix Q, by using change of variables, one obtains for every $f \in \mathcal{C}_L$ on $\mathbb{R}^d$,

$$\mathrm{Ent}_\nu(f^2) \leq 2 \int_{\mathbb{R}^d} (Q \nabla f) \nabla f \, d\nu. \tag{B.1}$$

One notes that the scheme (3.2) shows that for any $n \in \mathbb{N}$ and $x \in \mathbb{R}^d$, conditional on the previous step $\overline{X}_{n-1} = x$, $\overline{X}_n$ is a Gaussian random variable with mean $\mu(x) = x + \mu_\gamma(x)\gamma$

where
$$\mu_\gamma(x) = -\nabla U_\gamma(x) + \frac{\gamma}{2}\left(\left(\nabla^2 U \nabla U\right)_\gamma(x) - \vec{\Delta}(\nabla U)_\gamma(x)\right),$$
and covariance matrix $Q(x) = 2\gamma\left(\mathbf{I}_d - \gamma\nabla^2 U_\gamma(x) + \frac{\gamma^2}{3}(\nabla^2 U_\gamma(x))^2\right)$. Then, by using (B.1), one obtains
$$\mathrm{Ent}_\nu(f^2) \leq 2\int_{\mathbb{R}^d}(Q\nabla f)\nabla f\, d\nu \leq 2\int_{\mathbb{R}^d}\frac{14}{3}\gamma|\nabla f|^2\, d\nu.$$
Therefore, applying Proposition B.2 with $s = a$, $f = \sqrt{1+|x|^2}$ and $C = \frac{14}{3}\gamma$ yields the desired result, i.e.
$$R_\gamma V_a(x) = \mathbb{E}_x(V_a(\overline{X}_1)) \leq e^{\frac{7}{3}\gamma a^2}\exp\left\{a\mathbb{E}((1+|\overline{X}_1|^2)^{1/2}|\overline{X}_0 = x)\right\}.$$

B.3 Proof of inequality (3.27) in Theorem 3.4

To obtain (3.27), one consider the following cases

(i) If $m > \frac{7}{3}c^2$,
$$\begin{aligned}
&C\gamma^{3+\alpha}\mathbb{E}\left[V_c(\bar{x}_0)\right]\sum_{k=0}^n e^{-\frac{7}{3}c^2\gamma k - m\gamma(n-k)}\\
&= C\gamma^{3+\alpha}e^{-m\gamma n}\mathbb{E}\left[V_c(\bar{x}_0)\right]\sum_{k=0}^n e^{-\frac{7}{3}c^2\gamma k + m\gamma k}\\
&= C\gamma^{3+\alpha}e^{-m\gamma n}\mathbb{E}\left[V_c(\bar{x}_0)\right]\frac{e^{(n+1)(m-\frac{7}{3}c^2)\gamma}-1}{e^{(m-\frac{7}{3}c^2)\gamma}-1}\\
&\leq C\gamma^{3+\alpha}e^{-m\gamma n}\mathbb{E}\left[V_c(\bar{x}_0)\right]\frac{e^{n(m-\frac{7}{3}c^2)\gamma}}{1-e^{-(m-\frac{7}{3}c^2)\gamma}}\\
&\leq \frac{C\mathbb{E}\left[V_c(\bar{x}_0)\right]}{m-\frac{7}{3}c^2}e^{m\gamma}\gamma^{2+\alpha}e^{-\frac{7}{3}c^2(n+1)\gamma}.
\end{aligned}$$

(ii) For the case $m < \frac{7}{3}c^2$, we have
$$\begin{aligned}
&C\gamma^{3+\alpha}e^{-m\gamma n}\mathbb{E}\left[V_c(\bar{x}_0)\right]\sum_{k=0}^n e^{-\frac{7}{3}c^2\gamma k + m\gamma k}\\
&\leq C\gamma^{3+\alpha}e^{-m\gamma n}\mathbb{E}\left[V_c(\bar{x}_0)\right]\frac{1}{1-e^{-(\frac{7}{3}c^2-m)\gamma}}\\
&\leq \frac{C\mathbb{E}\left[V_c(\bar{x}_0)\right]}{\frac{7}{3}c^2-m}e^{\frac{7}{3}c^2\gamma}\gamma^{2+\alpha}e^{-m(n+1)\gamma}.
\end{aligned}$$

(iii) As for the case $m = \frac{7}{3}c^2$, it can be shown that
$$\begin{aligned}
&C\gamma^{3+\alpha}e^{-m\gamma n}\mathbb{E}\left[V_c(\bar{x}_0)\right]\sum_{k=0}^n e^{-\frac{7}{3}c^2\gamma k + m\gamma k}\\
&= C(n+1)\gamma^{3+\alpha}e^{-m\gamma n}\mathbb{E}\left[V_c(\bar{x}_0)\right]\\
&\leq \frac{C\mathbb{E}\left[V_c(\bar{x}_0)\right]}{m}e^{m\gamma}\gamma^{2+\alpha}.
\end{aligned}$$

B.4 Proof of inequality (3.46) in Lemma 3.23

For all $x, y \in \mathbb{R}^d$ and a constant $c > 0$, denote by $g(t) = \nabla^2 U(x + tc(y-x))$. One notes that for any $i, j = 1, \ldots, d$, $(g^{(i,j)})'(t) = c\sum_{k=1}^d \frac{\partial^3 U(x+tc(y-x))}{\partial x^{(i)}\partial x^{(j)}\partial x^{(k)}}(y^{(k)} - x^{(k)})$. By mean value theorem, there exists $t_{ij} \in [0,1]$, such that

$$\nabla^2 U^{(i,j)}(x + c(y-x)) - \nabla^2 U^{(i,j)}(x) = g^{(i,j)}(1) - g^{(i,j)}(0) = (g^{(i,j)})'(t_{ij}).$$

Then, one obtains

$$\begin{aligned}
&|\nabla^2 U(x + c(y-x)) - \nabla^2 U(x)|_{\mathsf{F}} \\
&= |g(1) - g(0)|_{\mathsf{F}} \\
&= c\sqrt{\sum_{i,j=1}^{d} \left| \sum_{k=1}^{d} \frac{\partial^3 U(x + t_{ij}c(y-x))}{\partial x^{(i)} \partial x^{(j)} \partial x^{(k)}} (y^{(k)} - x^{(k)}) \right|^2} \\
&\leq \sqrt{d} L_2 |c(y-x)|,
\end{aligned}$$

which, by sending c to zero yields

$$\sqrt{\sum_{i,j=1}^{d} \left| \sum_{k=1}^{d} \frac{\partial^3 U(x)}{\partial x^{(i)} \partial x^{(j)} \partial x^{(k)}} (y^{(k)} - x^{(k)}) \right|^2} \leq \sqrt{d} L_2 |y - x|.$$

B.5 Proof of inequality (3.46) in Lemma 3.23

For any $x \in \mathbb{R}^d$, our goal is to find an upper bound for

$$\sum_{i,j=1}^{d} \left| \sum_{k=1}^{d} \frac{\partial^4 U(x)}{\partial x^{(i)} \partial x^{(j)} \partial x^{(k)} \partial x^{(k)}} \right|^2 \leq d \sum_{k=1}^{d} \sum_{i,j=1}^{d} \left| \frac{\partial^4 U(x)}{\partial x^{(i)} \partial x^{(j)} \partial x^{(k)} \partial x^{(k)}} \right|^2.$$

For any $i,j,k = 1,\dots,d$, for all $x, y \in \mathbb{R}^d$ and a constant $c > 0$, define a function $g : \mathbb{R} \to \mathbb{R}^d$ by

$$g_{(i,j)}^{(k)}(t) = (\nabla(\nabla^2 U)^{(i,j)}(x + tcy))^{(k)} = \frac{\partial^3 U(x + tcy)}{\partial x^{(i)} \partial x^{(j)} \partial x^{(k)}}.$$

One notes that by mean value theorem, there exists $t_k \in [0,1]$, such that

$$\begin{aligned}
g_{(i,j)}^{(k)}(1) - g_{(i,j)}^{(k)}(0) &= (\nabla(\nabla^2 U)^{(i,j)}(x + cy))^{(k)} - (\nabla(\nabla^2 U)^{(i,j)}(x))^{(k)} \\
&= c \sum_{l=1}^{d} (\nabla^2(\nabla^2 U)^{(i,j)}(x + t_k cy))^{(k,l)} y^{(l)}.
\end{aligned}$$

Then, since

$$\begin{aligned}
&\left| \nabla((\nabla^2 U)^{(i,j)}(x + cy)) - \nabla((\nabla^2 U)^{(i,j)}(x)) \right| \\
&= \left(\sum_{k=1}^{d} \left| (\nabla(\nabla^2 U)^{(i,j)}(x + cy))^{(k)} - (\nabla(\nabla^2 U)^{(i,j)}(x))^{(k)} \right|^2 \right)^{1/2} \\
&= c \left(\sum_{k=1}^{d} \left| \sum_{l=1}^{d} (\nabla^2(\nabla^2 U)^{(i,j)}(x + t_k cy))^{(k,l)} y^{(l)} \right|^2 \right)^{1/2} \\
&= \left(\sum_{k=1}^{d} \left| \frac{\partial^3 U(x + cy)}{\partial x^{(i)} \partial x^{(j)} \partial x^{(k)}} - \frac{\partial^3 U(x)}{\partial x^{(i)} \partial x^{(j)} \partial x^{(k)}} \right|^2 \right)^{1/2} \\
&\leq \left| \nabla^2(\nabla U)^{(i)}(x + cy) - \nabla^2(\nabla U)^{(i)}(x) \right|_{\mathsf{F}} \\
&\leq \sqrt{d} L c |y|,
\end{aligned}$$

one obtains for any $i, j = 1, \ldots, d$ and $x \in \mathbb{R}^d$,

$$\left(\sum_{k=1}^{d} \left| \sum_{l=1}^{d} (\nabla^2 (\nabla^2 U)^{(i,j)} (x + t_k c y))^{(k,l)} y^{(l)} \right|^2 \right)^{1/2} \leq \sqrt{d} L |y|,$$

which, by sending c to zero yields

$$\left| \nabla^2 (\nabla^2 U)^{(i,j)} (x) y \right| \leq \sqrt{d} L |y|$$

and this implies $\left| \nabla^2 (\nabla^2 U)^{(i,j)} (x) \right| \leq \sqrt{d} L$. Finally, we have for any $x \in \mathbb{R}^d$,

$$\begin{aligned} d \sum_{k=1}^{d} \sum_{i,j=1}^{d} \left| \frac{\partial^4 U(x)}{\partial x^{(i)} \partial x^{(j)} \partial x^{(k)} \partial x^{(k)}} \right|^2 &\leq d \sum_{k=1}^{d} \left| \nabla^2 (\nabla^2 U)^{(k,k)} (x) \right|_{\mathsf{F}}^2 \\ &\leq d^2 \sum_{k=1}^{d} \left| \nabla^2 (\nabla^2 U)^{(k,k)} (x) \right|^2 \leq d^4 L^2. \end{aligned}$$

Appendix C

Auxiliary results to Chapter 4

C.1 Proof of the claim in Remark 4.3

We adapt the proof from [8, Lemma 4.7] and extend it to an $\mathbb{R}^m$-valued random variable Z_0. It suffices to consider $H(\theta, Z_0) = \dot{g}(\theta, Z_0)\mathbb{1}_{\bigcap_{i=1}^m \{Z_0^{(i)} \in I_i(\theta)\}}$, where $\theta \in \mathbb{R}^d$, $\dot{g}$ is bounded and jointly Lipschitz continuous, i.e. there exist $L_3, L_4, K_2 > 0$ such that for any $\theta, \theta' \in \mathbb{R}^d$, $z, z' \in \mathbb{R}^m$,

$$|\dot{g}(\theta, z) - \dot{g}(\theta', z')| \leq (1 + |z| + |z'|)^\rho (L_3|\theta - \theta'| + L_4|z - z'|), \quad |\dot{g}(\theta, z)| \leq K_2,$$

and the intervals $I_i(\theta)$ take the form $(-\infty, \bar{g}^{(i)}(\theta))$ with $\bar{g}^{(i)}$ Lipschitz. One notices that the proof follows the same lines when $I_i(\theta)$ takes the form $(\bar{g}^{(i)}(\theta), \infty)$, $(\tilde{g}^{(i)}(\theta), \hat{g}^{(i)}(\theta))$ with $\bar{g}^{(i)}, \tilde{g}^{(i)}, \hat{g}^{(i)}$ Lipschitz. One writes,

$$\begin{aligned}
|H(\theta, Z_0) - H(\theta', Z_0)| &\leq \left| \dot{g}(\theta, Z_0)\mathbb{1}_{\bigcap_{i=1}^m \left\{Z_0^{(i)} < \bar{g}^{(i)}(\theta)\right\}} - \dot{g}(\theta', Z_0)\mathbb{1}_{\bigcap_{i=1}^m \left\{Z_0^{(i)} < \bar{g}^{(i)}(\theta')\right\}} \right| \\
&\leq \left| \dot{g}(\theta, Z_0)\mathbb{1}_{\bigcap_{i=1}^m \left\{Z_0^{(i)} < \bar{g}^{(i)}(\theta)\right\}} - \dot{g}(\theta', Z_0)\mathbb{1}_{\bigcap_{i=1}^m \left\{Z_0^{(i)} < \bar{g}^{(i)}(\theta)\right\}} \right| \\
&\quad + \left| \dot{g}(\theta', Z_0)\mathbb{1}_{\bigcap_{i=1}^m \left\{Z_0^{(i)} < \bar{g}^{(i)}(\theta)\right\}} - \dot{g}(\theta', Z_0)\mathbb{1}_{\bigcap_{i=1}^m \left\{Z_0^{(i)} < \bar{g}^{(i)}(\theta')\right\}} \right| \\
&\leq L_3(1 + 2|Z_0|)^\rho|\theta - \theta'| + K_2\mathbb{1}_{\bigcap_{i=1}^m \left\{Z_0^{(i)} \in [\bar{g}^{(i)}(\theta), \bar{g}^{(i)}(\theta'))\right\}},
\end{aligned}$$

where we assume without loss of generality $\bar{g}^{(i)}(\theta) \leq \bar{g}^{(i)}(\theta')$ for all $i = 1, \ldots, m$. By taking expectation on both sides and by using Cauchy-Schwarz inequality, one obtains

$$\begin{aligned}
&\mathbb{E}\left[|H(\theta, Z_0) - H(\theta', Z_0)|\right] \\
&\leq L_3\mathbb{E}[(1 + 2|Z_0|)^\rho]|\theta - \theta'| + K_2\mathbb{P}\left(\bigcap_{i=1}^m \left\{Z_0^{(i)} \in [\bar{g}^{(i)}(\theta), \bar{g}^{(i)}(\theta'))\right\}\right) \\
&\leq L_3\mathbb{E}[(1 + 2|Z_0|)^\rho]|\theta - \theta'| + K_2\int_{\bar{g}^{(m)}(\theta)}^{\bar{g}^{(m)}(\theta')} \cdots \int_{\bar{g}^{(1)}(\theta)}^{\bar{g}^{(1)}(\theta')} f_{Z_0}(z^{(1)}, \ldots, z^{(m)}) dz^{(1)} \cdots dz^{(m)} \\
&\leq L_3\mathbb{E}[(1 + 2|Z_0|)^\rho]|\theta - \theta'| + K_2\int_{\bar{g}^{(1)}(\theta)}^{\bar{g}^{(1)}(\theta')} f_{Z_0^{(1)}}(z^{(1)}) dz^{(1)} \\
&\leq L_3\mathbb{E}[(1 + 2|Z_0|)^\rho]|\theta - \theta'| + K_2K_3L_5|\theta - \theta'| \\
&\leq (L_3 + K_2K_3L_5)\mathbb{E}[(1 + 2|Z_0|)^\rho]|\theta - \theta'|,
\end{aligned}$$

where $f_{Z_0^{(i)}}$ denotes the marginal density function of $Z_0^{(i)}$, K_3 is an upper bound of $f_{Z_0^{(1)}}$ and L_5 is a Lipschitz constant for $\bar{g}^{(1)}$. Taking $L = L_3 + K_2K_3L_5$ completes the proof.

C.2 Proof of the claim in Remark 4.10

By **C-5**, one obtains, for $\theta \in \mathbb{R}^d$ and $z \in \mathbb{R}^m$,

$$\langle F(\theta, z) - F(0, z), \theta \rangle \geq \langle \theta, \hat{A}_1(z)\theta \rangle,$$

which implies

$$\begin{aligned}
\langle F(\theta, z), \theta \rangle &\geq \langle \theta, \hat{A}_1(z)\theta \rangle + \langle F(0, z), \theta \rangle \\
&\geq \langle \theta, \hat{A}_1(z)\theta \rangle - |F(0, z)||\theta| \\
&\geq \langle \theta, \hat{A}_1(z)\theta \rangle - \epsilon|\theta|^2 - (L_2(1 + |z|)^{\rho+1} + |F(0, 0)|)^2/(4\epsilon) \\
&\geq \langle \theta, \hat{A}_1^*(z)\theta \rangle - \hat{b}(z),
\end{aligned}$$

where the third inequality holds due to **C-1** and $ab < \epsilon a^2 + b^2/(4\epsilon)$, for any $a, b > 0$, $\epsilon > 0$, $\hat{A}_1^*(z) = \hat{A}_1(z) - \epsilon \mathbf{I}_d$ and $\hat{b}(z) = (L_2(1 + |z|)^{\rho+1} + |F(0, 0)|)^2/(4\epsilon)$.

C.3 Proof of an inequality in equation (4.44)

The third inequality in (4.44) holds due to the following: for any $\omega \in \Omega$, one writes

$$\begin{aligned}
&\Big|(Z(\omega) - \bar{\theta}^i)\mathbb{1}_{\{-\sqrt{R}+\theta^i \leq Z(\omega) \leq \sqrt{R}+\theta^i\}}\mathbb{1}_{\mathcal{V}^i(\bar{\theta})}(Z(\omega)) \qquad\qquad (C.1)\\
&\quad -(Z(\omega) - \bar{\theta}^i)\mathbb{1}_{\{-\sqrt{R}+\bar{\theta}^i \leq Z(\omega) \leq \sqrt{R}+\bar{\theta}^i\}}\mathbb{1}_{\mathcal{V}^i(\bar{\theta})}(Z(\omega))\Big| \\
&\leq \Big|\mathbb{1}_{\{|Z(\omega)-\bar{\theta}^i| \leq \sqrt{R}\}} - \mathbb{1}_{\{|Z(\omega)-\theta^i| \leq \sqrt{R}\}}\Big| \, |Z(\omega) - \bar{\theta}^i| \qquad\qquad (C.2)\\
&= \begin{cases} 0, & \text{if} \quad \omega : |Z(\omega) - \bar{\theta}^i| \leq \sqrt{R}, \quad |Z(\omega) - \theta^i| \leq \sqrt{R}, \\ 0, & \text{if} \quad \omega : |Z(\omega) - \bar{\theta}^i| > \sqrt{R}, \quad |Z(\omega) - \theta^i| > \sqrt{R}, \\ \mathbb{1}_{\{|Z(\omega)-\bar{\theta}^i| \leq \sqrt{R}\}}\sqrt{R}, & \text{if} \quad \omega : |Z(\omega) - \bar{\theta}^i| \leq \sqrt{R}, \quad |Z(\omega) - \theta^i| > \sqrt{R}, \end{cases} \\
&\leq \mathbb{1}_{\{|Z(\omega)-\theta^i| \leq \sqrt{R}\}}(\sqrt{R} + |\theta^i - \bar{\theta}^i|), \quad \text{if} \quad \omega : |Z(\omega) - \bar{\theta}^i| > \sqrt{R}, \quad |Z(\omega) - \theta^i| \leq \sqrt{R},
\end{aligned}$$

and

$$\begin{aligned}
&\Big|\mathbb{1}_{\{|Z(\omega)-\bar{\theta}^i| \leq \sqrt{R}\}} - \mathbb{1}_{\{|Z(\omega)-\theta^i| \leq \sqrt{R}\}}\Big| (\sqrt{R} + |\theta^i - \bar{\theta}^i|) \qquad\qquad (C.3)\\
&= \begin{cases} 0, & \text{if} \quad \omega : |Z(\omega) - \bar{\theta}^i| \leq \sqrt{R}, \quad |Z(\omega) - \theta^i| \leq \sqrt{R}, \\ 0, & \text{if} \quad \omega : |Z(\omega) - \bar{\theta}^i| > \sqrt{R}, \quad |Z(\omega) - \theta^i| > \sqrt{R}, \\ \mathbb{1}_{\{|Z(\omega)-\bar{\theta}^i| \leq \sqrt{R}\}}(\sqrt{R} + |\theta^i - \bar{\theta}^i|), & \text{if} \quad \omega : |Z(\omega) - \bar{\theta}^i| \leq \sqrt{R}, \quad |Z(\omega) - \theta^i| > \sqrt{R}, \\ \mathbb{1}_{\{|Z(\omega)-\theta^i| \leq \sqrt{R}\}}(\sqrt{R} + |\theta^i - \bar{\theta}^i|), & \text{if} \quad \omega : |Z(\omega) - \bar{\theta}^i| > \sqrt{R}, \quad |Z(\omega) - \theta^i| \leq \sqrt{R}, \end{cases}
\end{aligned}$$

then one notices that (C.3) dominates (C.1).

C.4 Auxiliary results

Lemma C.1. *Assume **C-1**, **C-2**, **C-3** and **C-4** hold. For any $t \in [nT, (n+1)T]$, $n \in \mathbb{N}$ and $k = 1, \ldots, K+1$, $K+1 \leq T$, one obtains*

$$\mathbb{E}\left[\left|H(\bar{\theta}^\gamma_{nT+k-1}, Z_{nT+k}) - h(\bar{\theta}^\gamma_{nT+k-1})\right|^2\right] \leq e^{-a\gamma nT}\bar{\sigma}_Z \mathbb{E}[V_2(\theta_0)] + \tilde{\sigma}_Z,$$

where

$$\begin{aligned}
\bar{\sigma}_Z &= 4\mathbb{E}\left[K_\rho(Z_0)\right]\left(L^2 + L_1^2\right) \\
\tilde{\sigma}_Z &= 4\mathbb{E}\left[K_\rho(Z_0)\right]\left(L^2 + L_1^2\right)c_1(\gamma_{\max} + a^{-1}) + 4|h(0)|^2 + 8L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 8\mathbb{E}\left[F_*^2(Z_0)\right].
\end{aligned} \qquad (C.4)$$

Proof. One notices that by Remark 4.1 and 4.2,

$$\begin{aligned}
&\mathbb{E}\left[\left|H(\bar{\theta}^{\gamma}_{nT+k-1}, Z_{nT+k}) - h(\bar{\theta}^{\gamma}_{nT+k-1})\right|^2\right] \\
&\leq 2\mathbb{E}\left[\left|h(\bar{\theta}^{\gamma}_{nT+k-1})\right|^2\right] + 2\mathbb{E}\left[\left|H(\bar{\theta}^{\gamma}_{nT+k-1}, Z_{nT+k})\right|^2\right] \\
&\leq 2\mathbb{E}\left[\left(L\left|\bar{\theta}^{\gamma}_{nT+k-1}\right| + |h(0)|\right)^2\right] \\
&\quad + 2\mathbb{E}\left[\left((1+|Z_{nT+k}|)^{\rho+1}\left(L_1\left|\bar{\theta}^{\gamma}_{nT+k-1}\right| + L_2\right) + F_*(Z_{nT+k})\right)^2\right] \\
&\leq 4L^2\mathbb{E}\left[\left|\bar{\theta}^{\gamma}_{nT+k-1}\right|^2\right] + 4|h(0)|^2 + 4L_1^2\mathbb{E}\left[K_\rho(Z_0)\right]\mathbb{E}\left[\left|\bar{\theta}^{\gamma}_{nT+k-1}\right|^2\right] \\
&\quad + 8L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 8\mathbb{E}\left[F_*^2(Z_0)\right] \\
&\leq 4\mathbb{E}\left[K_\rho(Z_0)\right]\left(L^2 + L_1^2\right)\left(e^{-a\gamma nT}\mathbb{E}[V_2(\theta_0)] + c_1(\gamma_{\max} + a^{-1})\right) \\
&\quad + 4|h(0)|^2 + 8L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 8\mathbb{E}\left[F_*^2(Z_0)\right],
\end{aligned}$$

where the last inequality holdes due to Lemma 4.18. Finally, one obtains

$$\mathbb{E}\left[\left|h(\bar{Y}^{\gamma,n}_t) - H(\bar{Y}^{\gamma,n}_t, Z_{nT+k})\right|^2\right] \leq e^{-a\gamma nT}\bar{\sigma}_Z\mathbb{E}[V_2(\theta_0)] + \tilde{\sigma}_Z,$$

where $\bar{\sigma}_Z = 4\mathbb{E}\left[K_\rho(Z_0)\right]\left(L^2 + L_1^2\right)$ and $\tilde{\sigma}_Z = 4\mathbb{E}\left[K_\rho(Z_0)\right]\left(L^2 + L_1^2\right)c_1(\gamma_{\max} + a^{-1}) + 4|h(0)|^2 + 8L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 8\mathbb{E}\left[F_*^2(Z_0)\right]$. □

Lemma C.2. *Assume **C-1**, **C-2** and **C-4** hold. For any $t > 0$, one obtains*

$$\mathbb{E}\left[\left|\bar{\theta}^{\gamma}_t - \bar{\theta}^{\gamma}_{\lfloor t\rfloor}\right|^2\right] \leq \gamma(e^{-a\gamma\lfloor t\rfloor}\bar{\sigma}_Y\mathbb{E}[V_2(\theta_0)] + \tilde{\sigma}_Y),$$

where

$$\begin{aligned}
\bar{\sigma}_Y &= 2\gamma_{\max}L_1^2\mathbb{E}\left[K_\rho(Z_0)\right] \\
\tilde{\sigma}_Y &= 2\gamma_{\max}L_1^2\mathbb{E}\left[K_\rho(Z_0)\right]c_1(\gamma_{\max} + a^{-1}) + 4\gamma_{\max}L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 4\gamma_{\max}\mathbb{E}\left[F_*^2(Z_0)\right] + 2d\beta^{-1}.
\end{aligned} \tag{C.5}$$

Proof. For any $t > 0$, one calculates

$$\begin{aligned}
\mathbb{E}\left[\left|\bar{\theta}^{\gamma}_t - \bar{\theta}^{\gamma}_{\lfloor t\rfloor}\right|^2\right] &= \mathbb{E}\left[\left|-\gamma\int_{\lfloor t\rfloor}^t H(\bar{\theta}^{\gamma}_{\lfloor t\rfloor}, Z_{\lceil t\rceil})ds + \sqrt{2\beta^{-1}\gamma}(\tilde{w}^{\gamma}_t - \tilde{w}^{\gamma}_{\lfloor t\rfloor})\right|^2\right] \\
&\leq \gamma^2\mathbb{E}\left[\left((1+|Z_{\lceil t\rceil}|)^{\rho+1}(L_1|\bar{\theta}^{\gamma}_{\lfloor t\rfloor}| + L_2) + F_*(Z_{\lceil t\rceil})\right)^2\right] + 2d\gamma\beta^{-1},
\end{aligned}$$

where the inequality holds due to Remark 4.1 and by applying Lemma 4.18, one obtains

$$\begin{aligned}
\mathbb{E}\left[\left|\bar{\theta}^{\gamma}_t - \bar{\theta}^{\gamma}_{\lfloor t\rfloor}\right|^2\right] &\leq 2\gamma^2L_1^2\mathbb{E}\left[K_\rho(Z_0)\right]\mathbb{E}[|\bar{\theta}^{\gamma}_{\lfloor t\rfloor}|^2] + 4\gamma^2L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 4\gamma^2\mathbb{E}\left[F_*^2(Z_0)\right] + 2d\gamma\beta^{-1} \\
&\leq \gamma((1-a\gamma)^{\lfloor t\rfloor}\bar{\sigma}_Y\mathbb{E}[V_2(\theta_0)] + \tilde{\sigma}_Y),
\end{aligned}$$

where

$$\begin{aligned}
\bar{\sigma}_Y &= 2\gamma_{\max}L_1^2\mathbb{E}\left[K_\rho(Z_0)\right], \\
\tilde{\sigma}_Y &= 2\gamma_{\max}L_1^2\mathbb{E}\left[K_\rho(Z_0)\right]c_1(\gamma_{\max} + a^{-1}) + 4\gamma_{\max}L_2^2\mathbb{E}\left[K_\rho(Z_0)\right] + 4\gamma_{\max}\mathbb{E}\left[F_*^2(Z_0)\right] + 2d\beta^{-1}.
\end{aligned}$$

□

Lemma C.3. *Assume **C-1**, **C-2**, and **C-4** hold. Then, for any $t > 0$, one obtains*

$$\mathbb{E}[|\hat{Y}_t|^2] \le e^{-at}\mathbb{E}[|\theta_0|^2] + \left(\frac{2d}{a\beta} + \frac{2b}{a} + \frac{\mathbb{E}[K_1^2(Z_0)]}{a^2}\right)(1 - e^{-at}).$$

Proof. For any $t > 0$, by applying Itô's formula to $e^{at}|\hat{Y}_t|^2$, one obtains, almost surely

$$de^{at}|\hat{Y}_t|^2 = ae^{at}|\hat{Y}_t|^2 dt - 2e^{at}\langle \hat{Y}_t, h(\hat{Y}_t)\rangle dt + 2e^{at}\langle \hat{Y}_t, \sqrt{2\beta^{-1}}dB_t\rangle + 2d\beta^{-1}e^{at}dt.$$

Then, integrating both sides and taking expectation yield

$$e^{at}\mathbb{E}[|\hat{Y}_t|^2] = \mathbb{E}[|\theta_0|^2] + a\int_0^t e^{as}\mathbb{E}[|\hat{Y}_s|^2]ds - 2\int_0^t e^{as}\mathbb{E}[\langle \hat{Y}_s, h(\hat{Y}_s)\rangle]ds + 2d\beta^{-1}\int_0^t e^{as}ds,$$

which implies by using **C-4**

$$\begin{aligned}
e^{at}\mathbb{E}[|\hat{Y}_t|^2] &= \mathbb{E}[|\theta_0|^2] + a\int_0^t e^{as}\mathbb{E}[|\hat{Y}_s|^2]ds - 2a\int_0^t e^{as}\mathbb{E}[|\hat{Y}_s|^2]ds + 2b\int_0^t e^{as}ds \\
&\quad + 2\int_0^t e^{as}\mathbb{E}[|\hat{Y}_s|]\mathbb{E}[K_1(Z_0)]ds + 2d\beta^{-1}\int_0^t e^{as}ds \\
&\le \mathbb{E}[|\theta_0|^2] + (2b + \mathbb{E}[K_1^2(Z_0)]/a + 2d\beta^{-1})(e^{at} - 1)/a.
\end{aligned}$$

Finally, one obtains

$$\mathbb{E}[|\hat{Y}_t|^2] \le e^{-at}\mathbb{E}[|\theta_0|^2] + (2b + \mathbb{E}[K_1^2(Z_0)]/a + 2d\beta^{-1})(1 - e^{-at})/a.$$

□

C.5 Validity of assumption C-3 for VaR-CVaR algorithm in Section 4.5.3

We aim to show assumption **C-3** is valid for H_{w_1}. To achieve this, it is enough to prove

(1) The inequality $|\hat{g}_{w_1}(w, Z) - \hat{g}_{w_1}(\bar{w}, Z)| \le 2|w_1 - \bar{w}_1|\sum_i |Z_i|$ holds, and

(2) the last inequality in (4.47) is satisfied.

To prove $|\hat{g}_{w_1}(w, Z) - \hat{g}_{w_1}(\bar{w}, Z)| \le 2|w_1 - \bar{w}_1|\sum_i |Z_i|$, recall that for every $j = 1, \ldots, n$, $i \neq j$,

$$\frac{\partial g_j(w)}{\partial w_j} = \frac{e^{w_j}(\sum_{l\neq j} e^{w_l})}{(\sum_{l=1}^n e^{w_l})^2}, \quad \frac{\partial g_i(w)}{\partial w_j} = -\frac{e^{w_i}e^{w_j}}{(\sum_{l=1}^n e^{w_l})^2}.$$

Then, one calculates

$$\begin{aligned}
&|\hat{g}_{w_1}(w, Z) - \hat{g}_{w_1}(\bar{w}, Z)| \\
&= \left|\sum_{i=1}^n \frac{\partial g_i(w)}{\partial w_1}Z_i - \sum_{i=1}^n \frac{\partial g_i(\bar{w})}{\partial w_1}Z_i\right| \\
&\le \left|\frac{\partial g_1(w)}{\partial w_1} - \frac{\partial g_1(\bar{w})}{\partial w_1}\right||Z_1| + \left|\sum_{i\neq 1}\frac{\partial g_i(w)}{\partial w_1}Z_i - \sum_{i\neq 1}\frac{\partial g_i(\bar{w})}{\partial w_1}Z_i\right| \\
&\le \left|\frac{e^{w_1}(\sum_{l\neq 1} e^{w_l})}{(\sum_{l=1}^n e^{w_l})^2} - \frac{e^{\bar{w}_1}(\sum_{l\neq 1} e^{w_l})}{(\sum_{l\neq 1} e^{w_l} + e^{\bar{w}_1})^2}\right||Z_1| + \sum_{i\neq 1}\left|\frac{e^{w_i}e^{\bar{w}_1}}{(\sum_{l\neq 1} e^{w_l} + e^{\bar{w}_1})^2} - \frac{e^{w_i}e^{w_1}}{(\sum_{l=1}^n e^{w_l})^2}\right||Z_i| \\
&= \frac{\sum_{l\neq 1} e^{w_l}}{(\sum_{l=1}^n e^{w_l})^2(\sum_{l\neq 1} e^{w_l} + e^{\bar{w}_1})^2}\left|\left(\sum_{l\neq 1} e^{w_l}\right)^2\left(e^{w_1} - e^{\bar{w}_1}\right) + e^{\bar{w}_1}e^{w_1}\left(e^{\bar{w}_1} - e^{w_1}\right)\right||Z_1|
\end{aligned}$$

$$+\sum_{i\neq 1}\frac{e^{w_i}}{(\sum_{l=1}^n e^{w_l})^2(\sum_{l\neq 1}e^{w_l}+e^{\bar{w}_1})^2}\left|\left(\sum_{l\neq 1}e^{w_l}\right)^2\left(e^{\bar{w}_1}-e^{w_1}\right)+e^{\bar{w}_1}e^{w_1}\left(e^{w_1}-e^{\bar{w}_1}\right)\right||Z_i|$$
$$\leq 2|w_1-\bar{w}_1|\sum_{i=1}^n|Z_i|,$$

where the last inequality holds due to $1-e^{-x}\leq x$ for all $x\geq 0$.

To prove the last inequality in (4.47) is satisfied, we assume without loss of generality $g_n(w)=\max\{g_2(w),\ldots,g_n(w)\}$. Then,

(i) For $\bar{w}_1\geq w_1$, one calculates

$$\mathbb{E}\left[\sum_i|Z_i|\left|\mathbb{1}_{\{\sum_{i=1}^n g_i(w)Z_i\geq\theta\}}-\mathbb{1}_{\{\sum_{i=1}^n g_i(\bar{w})Z_i\geq\theta\}}\right|\right]\leq I_1+I_2, \tag{C.6}$$

where

$$\begin{aligned}
I_1&=\mathbb{E}\left[\sum_i|Z_i|\left|\mathbb{1}_{\{\sum_{i=1}^n g_i(w)Z_i\geq\theta\}}-\mathbb{1}_{\{\sum_{l\neq 1}g_l(w)Z_l+g_1(\bar{w})Z_1\geq\theta\}}\right|\right],\\
I_2&=\mathbb{E}\left[\sum_i|Z_i|\left|\mathbb{1}_{\{\sum_{l\neq 1}g_l(w)Z_l+g_1(\bar{w})Z_1\geq\theta\}}-\mathbb{1}_{\{\sum_{l\neq 1,2}g_l(w)Z_l+g_1(\bar{w})Z_1+g_2(\bar{w})Z_2\geq\theta\}}\right|\right]\\
&\quad+\cdots\\
&\quad+\mathbb{E}\left[\sum_i|Z_i|\left|\mathbb{1}_{\{g_n(w)Z_n+\sum_{l\neq n}g_l(\bar{w})Z_l\geq\theta\}}-\mathbb{1}_{\{\sum_{i=1}^n g_i(\bar{w})Z_i\geq\theta\}}\right|\right].
\end{aligned}$$

To estimate I_1, one writes

$$\begin{aligned}
I_1&\leq\mathbb{E}\left[\sum_i|Z_i|\mathbb{1}_{\{(\theta-\sum_{l\neq n}g_l(w)Z_l)/g_n(w)\leq Z_n\leq(\theta-g_1(\bar{w})Z_1-\sum_{l\neq 1,n}g_l(w)Z_l)/g_n(w)\}}\right]\\
&\quad+\mathbb{E}\left[\sum_i|Z_i|\mathbb{1}_{\{(\theta-g_1(\bar{w})Z_1-\sum_{l\neq 1,n}g_l(w)Z_l)/g_n(w)\}\leq Z_n\leq(\theta-\sum_{l\neq n}g_l(w)Z_l)/g_n(w)\}}\right].
\end{aligned}$$

The first term on the RHS of the inequality above can be further estimated as

$$\begin{aligned}
&\mathbb{E}\left[\sum_i|Z_i|\mathbb{1}_{\{(\theta-\sum_{l\neq n}g_l(w)Z_l)/g_n(w)\leq Z_n\leq(\theta-g_1(\bar{w})Z_1-\sum_{l\neq 1,n}g_l(w)Z_l)/g_n(w)\}}\right]\\
&=\mathbb{E}\left[\sum_{i\neq n}|Z_i|\mathbb{E}\left[\mathbb{1}_{\{(\theta-\sum_{l\neq n}g_l(w)Z_l)/g_n(w)\leq Z_n\leq(\theta-g_1(\bar{w})Z_1-\sum_{l\neq 1,n}g_l(w)Z_l)/g_n(w)\}}\,\middle|\,Z_1,\ldots,Z_{n-1}\right]\right]\\
&\quad+\mathbb{E}\left[\mathbb{E}\left[|Z_n|\mathbb{1}_{\{(\theta-\sum_{l\neq n}g_l(w)Z_l)/g_n(w)\leq Z_n\leq(\theta-g_1(\bar{w})Z_1-\sum_{l\neq 1,n}g_l(w)Z_l)/g_n(w)\}}\,\middle|\,Z_1,\ldots,Z_{n-1}\right]\right]\\
&=\int_{-\infty}^{\infty}\sum_{i\neq n}|z_i|\cdots\int_{-\infty}^{\infty}\int_{(\theta-\sum_{l\neq n}g_l(w)z_l)/g_n(w)}^{(\theta-g_1(\bar{w})z_1-\sum_{l\neq 1,n}g_l(w)z_l)/g_n(w)}f_{Z_n}(y)\,dy\\
&\qquad\qquad\times f_{Z_{n-1}}(z_{n-1})\,dz_{n-1}\cdots f_{Z_1}(z_1)\,dz_1\\
&\quad+\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\int_{(\theta-\sum_{l\neq n}g_l(w)z_l)/g_n(w)}^{(\theta-g_1(\bar{w})z_1-\sum_{l\neq 1,n}g_l(w)z_l)/g_n(w)}|z_n|f_{Z_n}(y)\,dy\\
&\qquad\qquad\times f_{Z_{n-1}}(z_{n-1})\,dz_{n-1}\cdots f_{Z_1}(z_1)\,dz_1\\
&\leq\frac{c_{Z_n}(c_{\bar{Z}}+(n-2)c_Z^2)}{g_n(w)}|g_1(w)-g_1(\bar{w})|+\frac{\bar{c}_{Z_n}c_Z}{g_n(w)}|g_1(w)-g_1(\bar{w})|
\end{aligned}$$

$$
\begin{aligned}
&= (c_{Z_n}(c_{\bar{Z}} + (n-2)c_Z^2) + \bar{c}_{Z_n} c_Z) \frac{\sum_i e^{w_i}}{e^{w_n}} \frac{\left(\sum_{i\neq 1} e^{w_i}\right) |e^{\bar{w}_1} - e^{w_1}|}{\left(\sum_i e^{w_i}\right)\left(e^{\bar{w}_1} + \sum_{i\neq 1} e^{w_i}\right)} \\
&\leq (c_{Z_n}(c_{\bar{Z}} + (n-2)c_Z^2) + \bar{c}_{Z_n} c_Z) \frac{\sum_{i\neq 1} g_i(w)}{g_n(w)} \frac{e^{\bar{w}_1}}{\left(e^{\bar{w}_1} + \sum_{i\neq 1} e^{w_i}\right)} |\bar{w}_1 - w_1| \\
&\leq (c_{Z_n}(c_{\bar{Z}} + (n-2)c_Z^2) + \bar{c}_{Z_n} c_Z)(n-1) \frac{e^{\bar{w}_1}}{\left(e^{\bar{w}_1} + \sum_{i\neq 1} e^{w_i}\right)} |\bar{w}_1 - w_1| \\
&\leq (c_{Z_n}(c_{\bar{Z}} + (n-2)c_Z^2) + \bar{c}_{Z_n} c_Z)(n-1) |\bar{w}_1 - w_1|,
\end{aligned}
$$

where $c_{\bar{Z}}$ denotes the second absolute moment of Z_i's , c_{Z_n} is the upper bound of the density of Z_n, and we use $1 - e^{-x} \leq x$ for $x \geq 0$ in the third inequality. Moreover, I_2 can be upper bounded by

$$
\begin{aligned}
I_2 \leq\ & \mathbb{E}\left[\sum_i |Z_i| \mathbb{1}_{\{(\theta - \sum_{l\neq 1} g_l(w) Z_l)/g_1(\bar{w}) \leq Z_1 \leq (\theta - \sum_{l\neq 1,2} g_l(w) Z_l - g_2(\bar{w}) Z_2)/g_1(\bar{w})\}}\right] \\
&+ \mathbb{E}\left[\sum_i |Z_i| \mathbb{1}_{\{(\theta - \sum_{l\neq 1,2} g_l(w) Z_l - g_2(\bar{w}) Z_2)/g_1(\bar{w}) \leq Z_1 \leq (\theta - \sum_{l\neq 1} g_l(w) Z_l)/g_1(\bar{w})\}}\right] \\
&+ \cdots \\
&+ \mathbb{E}\left[\sum_i |Z_i| \mathbb{1}_{\{(\theta - g_n(w) Z_n - \sum_{l\neq 1,n} g_l(\bar{w}) Z_l)/g_1(\bar{w}) \leq Z_1 \leq (\theta - \sum_{l\neq 1} g_l(\bar{w}) Z_l)/g_1(\bar{w})\}}\right] \\
&+ \mathbb{E}\left[\sum_i |Z_i| \mathbb{1}_{\{(\theta - \sum_{l\neq 1} g_l(\bar{w}) Z_l)/g_1(\bar{w}) \leq Z_1 \leq (\theta - g_n(w) Z_n - \sum_{l\neq 1,n} g_l(\bar{w}) Z_l)/g_1(\bar{w})\}}\right].
\end{aligned}
$$

The first term on the RHS of the inequality above can be calculated as

$$
\begin{aligned}
&\mathbb{E}\left[\sum_i |Z_i| \mathbb{1}_{\{(\theta - \sum_{l\neq 1} g_l(w) Z_l)/g_1(\bar{w}) \leq Z_1 \leq (\theta - \sum_{l\neq 1,2} g_l(w) Z_l - g_2(\bar{w}) Z_2)/g_1(\bar{w})\}}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[|Z_1| \mathbb{1}_{\{(\theta - \sum_{l\neq 1} g_l(w) Z_l)/g_1(\bar{w}) \leq Z_1 \leq (\theta - \sum_{l\neq 1,2} g_l(w) Z_l - g_2(\bar{w}) Z_2)/g_1(\bar{w})\}} \,\middle|\, Z_2, \ldots, Z_n\right]\right] \\
&\quad + \mathbb{E}\left[\sum_{i\neq 1} |Z_i| \mathbb{E}\left[\mathbb{1}_{\{(\theta - \sum_{l\neq 1} g_l(w) Z_l)/g_1(\bar{w}) \leq Z_1 \leq (\theta - \sum_{l\neq 1,2} g_l(w) Z_l - g_2(\bar{w}) Z_2)/g_1(\bar{w})\}} \,\middle|\, Z_2, \ldots, Z_n\right]\right] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{(\theta - \sum_{l\neq 1} g_l(w) z_l)/g_1(\bar{w})}^{(\theta - \sum_{l\neq 1,2} g_l(w) z_l - g_2(\bar{w}) z_2)/g_1(\bar{w})} |y| f_{Z_1}(y)\, dy f_{Z_n}(z_n)\, dz_n \cdots f_{Z_2}(z_2)\, dz_2 \\
&\quad + \int_{-\infty}^{\infty} \sum_{i\neq 1} |z_i| \cdots \int_{-\infty}^{\infty} \int_{(\theta - \sum_{l\neq 1} g_l(w) z_l)/g_1(\bar{w})}^{(\theta - \sum_{l\neq 1,2} g_l(w) z_l - g_2(\bar{w}) z_2)/g_1(\bar{w})} f_{Z_1}(y)\, dy \\
&\qquad\qquad \times f_{Z_n}(z_n)\, dz_n \cdots f_{Z_2}(z_2)\, dz_2 \\
&\leq \frac{\bar{c}_{Z_1} c_Z}{g_1(\bar{w})} |g_2(w) - g_2(\bar{w})| + \frac{c_{Z_1}(c_{\bar{Z}} + (n-2)c_Z^2)}{g_1(\bar{w})} |g_2(w) - g_2(\bar{w})| \\
&= (\bar{c}_{Z_1} c_Z + c_{Z_1}(c_{\bar{Z}} + (n-2)c_Z^2)) \frac{\left(e^{\bar{w}_1} + \sum_{i\neq 1} e^{w_i}\right)}{e^{\bar{w}_1}} \frac{e^{w_2} |e^{\bar{w}_1} - e^{w_1}|}{\left(\sum_i e^{w_i}\right)\left(e^{\bar{w}_1} + \sum_{i\neq 1} e^{w_i}\right)} \\
&\leq (\bar{c}_{Z_1} c_Z + c_{Z_1}(c_{\bar{Z}} + (n-2)c_Z^2)) \frac{e^{w_2} e^{\bar{w}_1}}{e^{\bar{w}_1}\left(\sum_i e^{w_i}\right)} |\bar{w}_1 - w_1| \\
&\leq (\bar{c}_{Z_1} c_Z + c_{Z_1}(c_{\bar{Z}} + (n-2)c_Z^2)) |\bar{w}_1 - w_1|,
\end{aligned}
$$

where c_Z denotes the first absolute moment of Z_i's and $\bar{c}_{Z_1}$ is the upper bound of the function $|z|f_{Z_1}$. Thus, in the case $\bar{w}_1 \geq w_1$, (C.6) becomes

$$\begin{aligned}
&\mathbb{E}\left[\sum_i |Z_i| \left| \mathbb{1}_{\{\sum_{i=1}^n g_i(w)Z_i \geq \theta\}} - \mathbb{1}_{\{\sum_{i=1}^n g_i(\bar{w})Z_i \geq \theta\}} \right|\right] \\
&\leq 2(n-1)((c_{Z_n} + c_{Z_1})(c_{\bar{Z}} + (n-2)c_Z^2) + c_Z(\bar{c}_{Z_n} + \bar{c}_{Z_1}))|\bar{w}_1 - w_1|.
\end{aligned}$$

(ii) As for the case $w_1 > \bar{w}_1$, the calculations are close to the above, however, one considers a different splitting as follows

$$\mathbb{E}\left[\sum_i |Z_i| \left| \mathbb{1}_{\{\sum_{i=1}^n g_i(w)Z_i \geq \theta\}} - \mathbb{1}_{\{\sum_{i=1}^n g_i(\bar{w})Z_i \geq \theta\}} \right|\right] \leq T_1 + T_2, \tag{C.7}$$

where

$$\begin{aligned}
T_1 = &\mathbb{E}\left[\sum_i |Z_i| \left| \mathbb{1}_{\{\sum_{i=1}^n g_i(w)Z_i \geq \theta\}} - \mathbb{1}_{\{\sum_{l\neq n} g_l(w)Z_l + g_n(\bar{w})Z_n \geq \theta\}} \right|\right] \\
&+ \cdots \\
&+ \mathbb{E}\left[\sum_i |Z_i| \left| \mathbb{1}_{\{g_1(w)Z_1 + g_2(w)Z_2 + \sum_{l\neq 1,2} g_l(\bar{w})Z_l \geq \theta\}} - \mathbb{1}_{\{g_1(w)Z_1 + \sum_{l\neq 1} g_l(\bar{w})Z_l \geq \theta\}} \right|\right], \\
T_2 = &\mathbb{E}\left[\sum_i |Z_i| \left| \mathbb{1}_{\{g_1(w)Z_1 + \sum_{l\neq 1} g_l(\bar{w})Z_l \geq \theta\}} - \mathbb{1}_{\{\sum_{i=1}^n g_i(\bar{w})Z_i \geq \theta\}} \right|\right].
\end{aligned}$$

To estimate T_1, one calculates

$$\begin{aligned}
T_1 \leq &\mathbb{E}\left[\sum_i |Z_i| \mathbb{1}_{\{(\theta - \sum_{l\neq 1} g_l(w)Z_l)/g_1(w) \leq Z_1 \leq (\theta - g_n(\bar{w})Z_n - \sum_{l\neq 1,n} g_l(w)Z_l)/g_1(w)\}}\right] \\
&+ \mathbb{E}\left[\sum_i |Z_i| \mathbb{1}_{\{(\theta - g_n(\bar{w})Z_n - \sum_{l\neq 1,n} g_l(w)Z_l)/g_1(w) \leq Z_1 \leq (\theta - \sum_{l\neq 1} g_l(w)Z_l)/g_1(w)\}}\right] \\
&+ \cdots \\
&+ \mathbb{E}\left[\sum_i |Z_i| \mathbb{1}_{\{(\theta - \sum_{l\neq 1,2} g_l(\bar{w})Z_l - g_2(w)Z_2)/g_1(w) \leq Z_1 \leq (\theta - \sum_{l\neq 1} g_l(\bar{w})Z_l)/g_1(w)\}}\right] \\
&+ \mathbb{E}\left[\sum_i |Z_i| \mathbb{1}_{\{(\theta - \sum_{l\neq 1} g_l(\bar{w})Z_l)/g_1(w) \leq Z_1 \leq (\theta - \sum_{l\neq 1,2} g_l(\bar{w})Z_l - g_2(w)Z_2)/g_1(w)\}}\right].
\end{aligned}$$

The first term on the RHS of the inequality above can be further calculated as

$$\begin{aligned}
&\mathbb{E}\left[\sum_i |Z_i| \mathbb{1}_{\{(\theta - \sum_{l\neq 1} g_l(w)Z_l)/g_1(w) \leq Z_1 \leq (\theta - g_n(\bar{w})Z_n - \sum_{l\neq 1,n} g_l(w)Z_l)/g_1(w)\}}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[|Z_1| \mathbb{1}_{\{(\theta - \sum_{l\neq 1} g_l(w)Z_l)/g_1(w) \leq Z_1 \leq (\theta - g_n(\bar{w})Z_n - \sum_{l\neq 1,n} g_l(w)Z_l)/g_1(w)\}} \,\middle|\, Z_2, \ldots, Z_n\right]\right] \\
&\quad + \mathbb{E}\left[\sum_{i\neq 1} |Z_i| \mathbb{E}\left[\mathbb{1}_{\{(\theta - \sum_{l\neq 1} g_l(w)Z_l)/g_1(w) \leq Z_1 \leq (\theta - g_n(\bar{w})Z_n - \sum_{l\neq 1,n} g_l(w)Z_l)/g_1(w)\}} \,\middle|\, Z_2, \ldots, Z_n\right]\right] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{(\theta - \sum_{l\neq 1} g_l(w)z_l)/g_1(w)}^{(\theta - g_n(\bar{w})z_n - \sum_{l\neq 1,n} g_l(w)z_l)/g_1(w)} |y| f_{Z_1}(y)\, dy f_{Z_n}(z_n)\, dz_n \cdots f_{Z_2}(z_2)\, dz_2 \\
&\quad + \int_{-\infty}^{\infty} \sum_{i\neq 1} |z_i| \cdots \int_{-\infty}^{\infty} \int_{(\theta - \sum_{l\neq 1} g_l(w)z_l)/g_1(w)}^{(\theta - g_n(\bar{w})z_n - \sum_{l\neq 1,n} g_l(w)z_l)/g_1(w)} f_{Z_1}(y)\, dy \\
&\qquad \times f_{Z_n}(z_n)\, dz_n \cdots f_{Z_2}(z_2)\, dz_2
\end{aligned}$$

$$
\begin{aligned}
&\le \frac{\bar{c}_{Z_1} c_Z}{g_1(w)} |g_n(w) - g_n(\bar{w})| + \frac{c_{Z_1}(c_{\bar{Z}} + (n-2)c_Z^2)}{g_1(w)} |g_n(w) - g_n(\bar{w})| \\
&= (\bar{c}_{Z_1} c_Z + c_{Z_1}(c_{\bar{Z}} + (n-2)c_Z^2)) \frac{\sum_i e^{w_i}}{e^{w_1}} \frac{e^{w_n} |e^{w_1} - e^{\bar{w}_1}|}{(\sum_i e^{w_i}) \left(e^{\bar{w}_1} + \sum_{i \neq 1} e^{w_i}\right)} \\
&\le (\bar{c}_{Z_1} c_Z + c_{Z_1}(c_{\bar{Z}} + (n-2)c_Z^2) \frac{e^{w_n} e^{w_1}}{e^{w_1} \left(e^{\bar{w}_1} + \sum_{i \neq 1} e^{w_i}\right)} |w_1 - \bar{w}_1| \\
&\le (\bar{c}_{Z_1} c_Z + c_{Z_1}(c_{\bar{Z}} + (n-2)c_Z^2) |w_1 - \bar{w}_1|.
\end{aligned}
$$

In addition, T_2 can be estimated as

$$
\begin{aligned}
T_2 \le\; &\mathbb{E}\left[\sum_i |Z_i| \mathbb{1}_{\{(\theta - g_1(w)Z_1 - \sum_{l \neq 1,n} g_l(\bar{w})Z_l)/g_n(\bar{w}) \le Z_n \le (\theta - \sum_{l \neq n} g_l(\bar{w})Z_l)/g_n(\bar{w})\}}\right] \\
&+ \mathbb{E}\left[\sum_i |Z_i| \mathbb{1}_{\{(\theta - \sum_{l \neq n} g_l(\bar{w})Z_l)/g_n(\bar{w}) \le Z_n \le (\theta - g_1(w)Z_1 - \sum_{l \neq 1,n} g_l(\bar{w})Z_l)/g_n(\bar{w})\}}\right].
\end{aligned}
$$

The first term on the RHS of the above inequality can be upper bounded by

$$
\begin{aligned}
&\mathbb{E}\left[\sum_i |Z_i| \mathbb{1}_{\{(\theta - g_1(w)Z_1 - \sum_{l \neq 1,n} g_l(\bar{w})Z_l)/g_n(\bar{w}) \le Z_n \le (\theta - \sum_{l \neq n} g_l(\bar{w})Z_l)/g_n(\bar{w})\}}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[|Z_n| \mathbb{1}_{\{(\theta - g_1(w)Z_1 - \sum_{l \neq 1,n} g_l(\bar{w})Z_l)/g_n(\bar{w}) \le Z_n \le (\theta - \sum_{l \neq n} g_l(\bar{w})Z_l)/g_n(\bar{w})\}} \,\middle|\, Z_1, \ldots, Z_{n-1}\right]\right] \\
&\quad + \mathbb{E}\left[\sum_{i \neq n} |Z_i| \mathbb{E}\left[\mathbb{1}_{\{(\theta - g_1(w)Z_1 - \sum_{l \neq 1,n} g_l(\bar{w})Z_l)/g_n(\bar{w}) \le Z_n \le (\theta - \sum_{l \neq n} g_l(\bar{w})Z_l)/g_n(\bar{w})\}} \,\middle|\, Z_1, \ldots, Z_{n-1}\right]\right] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{(\theta - g_1(w)z_1 - \sum_{l \neq 1,n} g_l(\bar{w})z_l)/g_n(\bar{w})}^{(\theta - \sum_{l \neq n} g_l(\bar{w})z_l)/g_n(\bar{w})} |z_n| f_{Z_n}(y)\, dy \\
&\qquad\qquad \times f_{Z_{n-1}}(z_{n-1})\, dz_{n-1} \cdots f_{Z_1}(z_1)\, dz_1 \\
&\quad + \int_{-\infty}^{\infty} \sum_{i \neq n} |z_i| \cdots \int_{-\infty}^{\infty} \int_{(\theta - g_1(w)z_1 - \sum_{l \neq 1,n} g_l(\bar{w})z_l)/g_n(\bar{w})}^{(\theta - \sum_{l \neq n} g_l(\bar{w})z_l)/g_n(\bar{w})} f_{Z_n}(y)\, dy \\
&\qquad\qquad \times f_{Z_{n-1}}(z_{n-1})\, dz_{n-1} \cdots f_{Z_1}(z_1)\, dz_1 \\
&\le \frac{\bar{c}_{Z_n} c_Z}{g_n(\bar{w})} |g_1(w) - g_1(\bar{w})| + \frac{c_{Z_n}(c_{\bar{Z}} + (n-2)c_Z^2)}{g_n(\bar{w})} |g_1(w) - g_1(\bar{w})| \\
&= (\bar{c}_{Z_n} c_Z + c_{Z_n}(c_{\bar{Z}} + (n-2)c_Z^2)) \frac{\left(e^{\bar{w}_1} + \sum_{i \neq 1} e^{w_i}\right)}{e^{w_n}} \frac{\left(\sum_{i \neq 1} e^{w_i}\right) |e^{w_1} - e^{\bar{w}_1}|}{(\sum_i e^{w_i}) \left(e^{\bar{w}_1} + \sum_{i \neq 1} e^{w_i}\right)} \\
&= (\bar{c}_{Z_n} c_Z + c_{Z_n}(c_{\bar{Z}} + (n-2)c_Z^2)) \frac{\sum_{i \neq 1} g_i(w)}{g_n(w)} \frac{e^{w_1}}{\sum_i e^{w_i}} |w_1 - \bar{w}_1| \\
&\le (n-1)(\bar{c}_{Z_n} c_Z + c_{Z_n}(c_{\bar{Z}} + (n-2)c_Z^2)) \frac{e^{w_1}}{\sum_i e^{w_i}} |w_1 - \bar{w}_1| \\
&\le (n-1)(\bar{c}_{Z_n} c_Z + c_{Z_n}(c_{\bar{Z}} + (n-2)c_Z^2)) |w_1 - \bar{w}_1|.
\end{aligned}
$$

Thus for the case $w_1 > \bar{w}_1$, we have

$$
\begin{aligned}
&\mathbb{E}\left[\sum_i |Z_i| \left|\mathbb{1}_{\{\sum_{i=1}^n g_i(w)Z_i \ge \theta\}} - \mathbb{1}_{\{\sum_{i=1}^n g_i(\bar{w})Z_i \ge \theta\}}\right|\right] \\
&\le 2(n-1)(c_Z(\bar{c}_{Z_n} + \bar{c}_{Z_1}) + (c_{\bar{Z}} + (n-2)c_Z^2)(c_{Z_n} + c_{Z_1})) |w_1 - \bar{w}_1|.
\end{aligned}
$$

Combining the two cases, one obtains

$$\begin{aligned}
&\mathbb{E}\left[\sum_i |Z_i| \left| \mathbb{1}_{\{\sum_{i=1}^n g_i(w) Z_i \geq \theta\}} - \mathbb{1}_{\{\sum_{i=1}^n g_i(\bar{w}) Z_i \geq \theta\}} \right|\right] \\
&\leq 2(n-1)(c_Z(\bar{c}_{Z_n} + \bar{c}_{Z_1}) + (c_{\bar{Z}} + (n-2)c_Z^2)(c_{Z_n} + c_{Z_1}))|w_1 - \bar{w}_1|.
\end{aligned}$$

Bibliography

[1] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*, volume 348. Springer Science & Business Media, 2013.

[2] Olivier Bardou, Noufel Frikha, and Gilles Pages. Computing VaR and CVaR using stochastic approximation and adaptive unconstrained importance sampling. *Monte Carlo Methods and Applications*, 15(3):173–210, 2009.

[3] M Barkhagen, NH Chau, É Moulines, M Rásonyi, S Sabanis, and Y Zhang. On stochastic gradient Langevin dynamics with dependent data streams in the logconcave case. *arXiv preprint arXiv:1812.02709*, 2018.

[4] Michel Benaïm, Jean-Claude Fort, and Gilles Pagès. Convergence of the one-dimensional Kohonen algorithm. *Advances in Applied Probability*, 30(3):850–869, 1998.

[5] Wolf-Jürgen Beyn, Elena Isaak, and Raphael Kruse. Stochastic C-stability and B-consistency of explicit and implicit Milstein-type schemes. *Journal of Scientific Computing*, 70(3):1042–1077, 2017.

[6] Nicolas Brosse, Alain Durmus, and Eric Moulines. The promises and pitfalls of Stochastic Gradient Langevin Dynamics. In *Advances in Neural Information Processing Systems*, pages 8268–8278, 2018.

[7] Nicolas Brosse, Alain Durmus, Éric Moulines, and Sotirios Sabanis. The Tamed Unadjusted Langevin Algorithm. *Stochastic Processes and their Applications*, 129(10):3638–3663, 2019.

[8] Huy N Chau, Chaman Kumar, Miklós Rásonyi, and Sotirios Sabanis. On fixed gain recursive estimators with discontinuity in the parameters. *ESAIM: Probability and Statistics*, 23:217–244, 2019.

[9] Ngoc Huy Chau, Éric Moulines, Miklos Rásonyi, Sotirios Sabanis, and Ying Zhang. On stochastic gradient Langevin dynamics with dependent data streams: the fully non-convex case. *arXiv preprint arXiv:1905.13142*, 2019.

[10] Xiang Cheng, Niladri S Chatterji, Yasin Abbasi-Yadkori, Peter L Bartlett, and Michael I Jordan. Sharp convergence rates for Langevin dynamics in the nonconvex setting. *arXiv preprint arXiv:1805.01648*, 2018.

[11] Arnak S Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):651–676, 2017.

[12] Arnak S Dalalyan and Avetik Karagulyan. User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient. *Stochastic Processes and their Applications*, 129(12):5278–5311, 2019.

[13] AM Davie. Pathwise approximation of stochastic differential equations using coupling. `www.maths.ed.ac.uk/~adavie/coum.pdf`, 2014.

[14] Alain Durmus and Eric Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *The Annals of Applied Probability*, 27(3):1551–1587, 2017.

[15] Alain Durmus and Eric Moulines. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Bernoulli*, 25(4A):2854–2882, 2019.

[16] Andreas Eberle. Reflection couplings and contraction rates for diffusions. *Probability theory and related fields*, 166(3-4):851–886, 2016.

[17] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Quantitative Harris-type theorems for diffusions and McKean–Vlasov processes. *Transactions of the American Mathematical Society*, 371(10):7135–7173, 2019.

[18] Gersende Fort, Éric Moulines, Amandine Schreck, and Matti Vihola. Convergence of Markovian Stochastic Approximation with discontinuous dynamics. *SIAM Journal on Control and Optimization*, 54(2):866–893, 2016.

[19] Andreas Griewank. Some bounds on the complexity of gradients, Jacobians, and Hessians. In *Complexity in numerical optimization*, pages 128–162. World Scientific, 1993.

[20] Andreas Griewank, Jean Utke, and Andrea Walther. Evaluating higher derivative tensors by forward propagation of univariate Taylor series. *Mathematics of Computation*, 69(231):1117–1130, 2000.

[21] Andreas Griewank and Andrea Walther. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, volume 105. Siam, 2008.

[22] István Gyöngy and Nicolai Krylov. On the rate of convergence of splitting-up approximations for SPDEs. In *Stochastic inequalities and applications*, pages 301–321. Springer, 2003.

[23] Patrik O Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of machine learning research*, 5(Nov):1457–1469, 2004.

[24] Martin Hutzenthaler, Arnulf Jentzen, and Peter E Kloeden. Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 467(2130):1563–1576, 2011.

[25] Martin Hutzenthaler, Arnulf Jentzen, and Peter E Kloeden. Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. *The Annals of Applied Probability*, 22(4):1611–1641, 2012.

[26] Chii-Ruey Hwang et al. Laplace's method revisited: weak convergence of probability measures. *The Annals of Probability*, 8(6):1177–1182, 1980.

[27] Peter E Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*, volume 23. Springer Science & Business Media, 2013.

[28] Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: Journal of the Econometric Society*, pages 33–50, 1978.

[29] Solomon Kullback. *Information Theory and Statistics*. Courier Corporation, 1997.

[30] Chaman Kumar and Sotirios Sabanis. On Milstein approximations with varying coefficients: the case of super-linear diffusion coefficients. *BIT Numerical Mathematics*, 59(4):929–968, 2019.

[31] Michel Ledoux. Concentration of measure and logarithmic Sobolev inequalities. In *Seminaire de probabilites XXXIII*, pages 120–216. Springer, 1999.

[32] Robert S Liptser and Albert N Shiryaev. *Statistics of Random Processes: I. General Theory*, volume 5. Springer Science & Business Media, 2013.

[33] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, 11(Jan):19–60, 2010.

[34] Mateusz B Majka, Aleksandar Mijatović, and Lukasz Szpruch. Non-asymptotic bounds for sampling algorithms without log-concavity. *arXiv preprint arXiv:1808.07105*, 2018.

[35] Xuerong Mao. *Stochastic Differential Equations and Applications*. Elsevier, 2007.

[36] Gisiro Maruyama. Continuous Markov processes and stochastic equations. *Rendiconti del Circolo Matematico di Palermo*, 4(1):48, 1955.

[37] Jonathan C Mattingly, Andrew M Stuart, and Desmond J Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic processes and their applications*, 101(2):185–232, 2002.

[38] Sean P Meyn and Richard L Tweedie. Stability of Markovian processes III: Foster–Lyapunov criteria for continuous-time processes. *Advances in Applied Probability*, 25(3):518–548, 1993.

[39] Sean P Meyn and Richard L Tweedie. *Markov Chains and Stochastic Stability*. Springer Science & Business Media, 2012.

[40] Yurii Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*, volume 87. Springer Science & Business Media, 2013.

[41] E. Platen and W. Wagner. On a Taylor formula for a class of Itô processes. *Probab. Math. Statist.*, 3(1):37–51, 1982.

[42] Maxim Raginsky, Alexander Rakhlin, and Matus Telgarsky. Non-convex learning via Stochastic Gradient Langevin Dynamics: a nonasymptotic analysis. In *Conference on Learning Theory*, pages 1674–1703, 2017.

[43] Gareth O Roberts and Richard L Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

[44] Sotirios Sabanis. A note on tamed Euler approximations. *Electronic Communications in Probability*, 18:1–10, 2013.

[45] Sotirios Sabanis. Euler approximations with varying coefficients: the case of superlinearly growing diffusion coefficients. *The Annals of Applied Probability*, 26(4):2083–2105, 2016.

[46] Sotirios Sabanis and Ying Zhang. Higher Order Langevin Monte Carlo algorithm. *Electronic Journal of Statistics*, 13(2):3805–3850, 2019.

[47] Sotirios Sabanis and Ying Zhang. On explicit order 1.5 approximations with varying coefficients: the case of super-linear diffusion coefficients. *Journal of Complexity*, 50:84–115, 2019.

[48] Sotirios Sabanis and Ying Zhang. A fully data-driven approach to minimizing CVaR for portfolio of assets via SGLD with discontinuous updating. *arXiv preprint*, 2020.

[49] Ichiro Takeuchi, Quoc V Le, Timothy D Sears, and Alexander J Smola. Nonparametric quantile estimation. *Journal of machine learning research*, 7(Jul):1231–1264, 2006.

[50] Xiaojie Wang and Siqing Gan. The tamed Milstein method for commutative stochastic differential equations with non-globally Lipschitz continuous coefficients. *Journal of Difference Equations and Applications*, 19(3):466–490, 2013.

[51] Max Welling and Yee W Teh. Bayesian Learning via Stochastic Gradient Langevin Dynamics. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 681–688, 2011.

[52] Pan Xu, Jinghui Chen, Difan Zou, and Quanquan Gu. Global convergence of Langevin dynamics based algorithms for nonconvex optimization. In *Advances in Neural Information Processing Systems*, pages 3122–3133, 2018.

[53] Ying Zhang, Ömer Deniz Akyildiz, Theo Damoulas, and Sotirios Sabanis. Nonasymptotic estimates for stochastic gradient langevin dynamics under local conditions in nonconvex optimization. *arXiv preprint*, 2020.